February
2026

Vol. 25
No. 1

Journal of Medical English Education
年3回2月・6月・10月発行　第25巻第1号　2026年2月1日発行
ISSN 1883-0951

Journal of Medical English Education

Japan Society for Medical English Education

Official Journal of the Japan Society for Medical English Education (JASMEE)

Journal of
Medical English Education
Official Journal of the Japan Society for Medical English Education (JASMEE)
Vol. 25 No. 1 February 2026 — Index

From the editor

1 **AI in the classroom** Alan Hauk

Original articles

2 **Ideal medical English textbook: From students' perspectives** Chie Saito

8 **Developing global competence in future healthcare professionals through a New Zealand study program (2009~2024)** Rie Sato, et al

17 **Enriching vocabulary learning through affixation: Classroom implications for medical English education** Haruko Shimazaki

23 **Refining the Japan Functional History Taking Assessment (JFHTA): A pilot study to optimize interrater reliability using PDCA cycles** Eric H. Jego, et al

Conference proceedings: The 28th JASMEE Academic Meeting（第 28 回日本医学英語教育学会学術集会）

30 **Health screening tourism: Toward improving hospitality for foreign tourists in Osaka** Tsuyoshi Sakoda, et al

Journal of Medical English Education

Vol. 25, No. 1, February 2026

Journal of Medical English Education, the official publication of The Japan Society for Medical English Education, was founded in 2000 to promote international exchange of knowledge in the field of English education for medical purposes. Until June 2006 (Vol. 5 No. 2), the registered title of the Journal was *Medical English - Journal of Medical English Education*; the current title, which was registered in December 2006 (Vol. 6 No. 1), should be used for citation purposes.

Copyright © 2026 by The Japan Society for Medical English Education
All rights reserved.

The Japan Society for Medical English Education
c/o NARUNIA Inc.
3–4–3 Hongo, Bunkyo–ku, Tokyo, 113–0033 Japan
TEL 03–3818–6450 (outside Japan: +81–3–3818–6450)
FAX 03–3818–0554 (outside Japan: +81–3–3818–0554)
E-MAIL jasmee@narunia.co.jp
WEBSITE https://jasmee.jp

Distributed by NARUNIA Inc.
3–4–3 Hongo, Bunkyo–ku, Tokyo, 113–0033 Japan

Journal of Medical English Education Vol. 25 No. 1
日本医学英語教育学会会誌
編集人 小島多香子
2026 年 2 月 1 日発行 第 25 巻 第 1 号
企画 日本医学英語教育学会
制作・発行 有限会社編集室なるにあ
〒 113-0033 東京都文京区本郷 3-4-3
TEL 03-3818-6450 ／ FAX 03-3818-0554 ／ E-MAIL jasmee@narunia.co.jp

Journal of Medical English Education

The Official Journal of the Japan Society for Medical English Education

jasmee@narunia.co.jp

Executive Chair, JASMEE Publications

Isao Date, Okayama

Editorial Committee

Editor-in-Chief

Takako Kojima, Tokyo

Associate Editor

Alan Hauk, Tokyo

Japanese Editor

Joji Tokugawa, Tokyo

Committee Members

Shinobu Hattori, Mie
James Hobbs, Tokyo
Eric Jego, Tokyo
Thomas Mayers, Ibaraki
Takayuki Oshimi, Chiba
Ian Willey, Kagawa

Executive Adviser

Timothy D. Minton, Tokyo

Editorial Executive Board

Raoul Breugelmans, Osaka
Isao Date, Okayama
Yoshitaka Fukuzawa, Aichi
Shinobu Hattori, Mie
Masahito Hitosugi, Shiga
Shigeo Irimajiri, Osaka
Jun Iwata, Shimane
Ikuo Kageyama, Niigata
Junichi Kameoka, Miyagi
Takako Kojima, Tokyo
Kazuhiko Kurozumi, Shizuoka
Timothy D. Minton, Tokyo
Shigeru Mori, Oita
Takayuki Oshimi, Chiba
Kazuaki Shimoji, Chiba
Kinko Tamamaki, Hyogo
Joji Tokugawa, Tokyo

Former Editors-in-Chief

Timothy D. Minton, 2014–2024
Reuben M. Gerling, 2008–2014
Nell L. Kennedy, 2004–2008
Shizuo Oi, 2000–2004

Executive Adviser Emeritus

Kenichi Uemura

From the editor

AI in the classroom

What should we do about our students using AI? This is a question that often comes up in conversations with other professors, the implication being that using AI is cheating and needs to be stopped. Until recently, I was in the "AI should not be allowed" group because I thought that students would not learn as much if they depended on AI to do their assignments. Furthermore, I felt that AI could not be trusted because it often "hallucinates," which is to say that if it does not know the answer to a question, then it just invents something that sounds good. I also realize, however, that AI is improving very quickly and will become an important tool for doctors, so it would be useful to start incorporating AI into my teaching.

This year, therefore, I encouraged my students to use AI for one particular assignment. I asked them to take a short passage about a disease from a medical text and explain it to me as if I had no understanding of medicine or medical terminology. AI-generated answers were permitted as long as they were accurate according to the text and easy for a layperson to understand. If an answer was wrong, hard to understand, or gave information outside of the text, I asked the student to refine their answer and submit it again. Since students will probably use AI in a similar way after they become doctors, I wanted them to try it out with me, first, and I was fairly pleased with the results.

Of course, in addition to worrying about accuracy, another fear of AI is that it will be used unethically. I asked a colleague about this, and his response was, "Let's ask ChatGPT." He opened the app on his phone and asked, "How can we teach our students to use AI responsibly?" I expected the answer to be something along the lines of, "Don't worry. AI is completely trustworthy and never lies or gives bad answers." To my surprise, though, ChatGPT's answer was very logical, emphasizing the need to verify all information, ask for citations and check them, and to use AI to help understand the flaws in your own work rather than to do the work for you from the beginning.

I followed this up with an Internet search which led me to an article on Harvard University's Graduate School of Education website that gave four key suggestions. First, do not ignore AI. It's here to stay, so learn to use it. Second, use AI in the classroom alongside your students. Look at AI-generated answers in class and discuss them. Third, teach students how to ask the right questions the right way. Ask follow-up questions to refine the query until AI knows exactly what you are asking. And, fourth, use AI to trigger creativity. This means that teachers must change their lessons so that exercises cannot be finished by AI alone. (https://www.gse.harvard.edu/ideas/usable-knowledge/23/07/embracing-artificial-intelligence-classroom)

So, while I may want to live in the comfortable past, I am going to be incorporating more AI in my teaching.

(Disclaimer: AI was not used either to write this article or to edit it, but perhaps it should have been.)

Alan Hauk
Associate Editor
Journal of Medical English Education

Original article

Ideal medical English textbook: From students' perspectives

Chie Saito
General Medical Education and Research Center, Teikyo University

Abstract

In 2024, Teikyo University School of Medicine introduced a joint first-year course integrating Anatomy and English to establish a strong foundation in medical English. Selecting an appropriate textbook for such a course is a continuing challenge, as materials must balance academic rigor, clinical relevance, and accessibility. This study aimed to evaluate two medical English textbooks, *English for Medical Students* (EMS) and *Because We Care* (BWC), from the perspective of first-year medical students. A questionnaire survey was administered to 91 students at the end of the course. The instrument included a five-point Likert scale rating across eight categories—content intelligibility, sentence readability, vocabulary, exercises, visual support, unit structure, relevance to the national medical examination and clinical practice, and overall satisfaction—as well as an open-ended section for free responses. Quantitative data were analyzed descriptively, and qualitative comments were coded thematically. Results showed that EMS achieved a higher overall satisfaction rating (M = 4.30) than BWC (M = 3.87). EMS was praised for its academic depth, particularly vocabulary and integration with anatomy, but criticized for outdated content and insufficient visual support. In contrast, BWC's strongest feature was visual clarity and listening tasks, which enhanced accessibility, yet students found it overly simple and questioned its usefulness for improving English proficiency. These findings suggest that medical English textbooks should combine the academic rigor of EMS with the communicative accessibility of BWC. Incorporating accurate content, clear visuals, and scaffolded output tasks may address student needs and inform the development of future medical English materials.

J Med Eng Educ (2026) 25 (1): 2-7

Keywords medical English, anatomy, ESP, CLIL, textbook evaluation, student perceptions

1. Introduction

In 2024, Teikyo University School of Medicine introduced a joint course integrating Anatomy and English as part of the first-year curriculum. While English is taught throughout all four years of the program, only the first-year course is delivered by language instructors like myself; from the second year onward, English instruction is provided by medical faculty members.

The joint course was designed to expose first-year students to Medical English from the outset and help them establish a solid foundation for future learning. In this collaboration, faculty members from both Anatomy and English courses participate in each other's classes. We selected a Medical English textbook containing anatomical reading passages, with the anatomy professor contributing by creating exam questions, reviewing content, and offering subject-matter expertise.

Corresponding author:
Chie Saito
General Medical Education and Research Center, Teikyo University,
2-11-1 Kaga, Itabashi-Ku, Tokyo 173-8605, Japan
TEL: 03-3964-1211
E-mail: saitochie@med.teikyo-u.ac.jp

1.1 Medical English textbooks

The following two medical English textbooks are currently used in the English component of this course. *English for Medical Students* is an academically oriented textbook that focuses on anatomical and physiological content. It comprises 20 chapters, primarily organized by body systems. Certain topics, such as the sensory organs and the brain, are presented in separate, dedicated chapters. Each chapter begins with an anatomical diagram drawn by a medical student. While these illustrations may lack the refinement of professionally created figures, some display notable detail. Beneath each diagram, the key terms from the illustration are listed.

The main section of each chapter consists of a reading passage of approximately 500–700 words, covering anatomical and physiological topics in an academic style.

Following the reading, there are two types of comprehension exercises: a fill-in-the-blank activity and a true-or-false exercise. The chapter concludes with a listening task in which students hear a conversation between a doctor and a patient. They then take notes on the chief complaint, onset and course of illness, diagnosis, and disposition. This is followed by explanations of selected medical English terms. The overall emphasis of this textbook is on developing reading proficiency and acquiring medical terminology.

Because We Care is a communicative textbook that emphasizes spoken interaction, listening comprehension, and the practical use of English in clinical contexts. It contains 14 units, each organized around a body system. Every unit begins with a short manga excerpt from Black Jack as a warm-up activity, followed by three conversational listening exercises, which can also be used as speaking practice.

Each listening task is accompanied by vocabulary exercises, including terms for symptoms typically associated with diseases of the relevant body system, disease names, and common abbreviations used in the field. The final section of each unit presents a relatively long reading passage, often narrating a patient case, followed by comprehension questions. Some listening and reading texts are repurposed for multiple exercises; for example, the end-of-unit reading passage is also used for a sequencing task in which students arrange illustrated scenes in the order they occur in the story.

This textbook offers a wide variety of tasks that develop listening and reading skills while building medical vocabulary. Its design accommodates learners at different proficiency levels and prepares students for a range of careers in healthcare and related fields.

1.2 Rationale for using both textbooks

Finding an appropriate medical English coursebook for medical students is a persistent challenge. Many available textbooks focus heavily on medical terminology. For example, *Medical English for Required Medical English* covers a vast range of terms from both macroscopic and microscopic perspectives and introduces useful phrases for medical settings, along with sample medical report forms. However, despite its valuable content, its overall format makes it difficult to adopt as the primary course textbook.

Another important factor is text difficulty. Even if the structure and format are well-suited for a course, a mismatch in difficulty level can reduce a textbook's effectiveness. For instance, *Introduction to Medical English* is exemplary in its overall organization and chapter design. It offers reading passages on body systems and diseases, substantial vocabulary exercises, and clear explanations. Nevertheless, some reading texts are overly simple for medical students, who are also studying demanding subjects such as anatomy and physiology. In such cases, materials that are too easy may not provide the level of challenge needed to support their academic growth.

This is not to suggest that these textbooks are ineffective or inappropriate for Medical English instruction. In fact, I have recommended *Medical English for Required Medical English* to some students for self-study, and I have successfully used *Introduction to Medical English* for other courses for many years. Rather, the point is that selecting a suitable textbook for medical students requires considering multiple factors, including content scope, format, difficulty, and relevance.

With these considerations in mind, we chose to use *English for Medical Students* (EMS) and *Because We Care* (BWC) in combination for our joint Anatomy–English course. EMS is more academic, reading-focused, and challenging, whereas BWC is more communicative, casual in tone, and comparatively easier. Their shared strength is that both are devoted entirely to medical English, ensuring high content relevance. By combining them, we aim to balance the development of all four macro-skills—reading, writing, listening, and speaking—while addressing the diverse needs and abilities of our students.

2. Methods

2.1 Purpose of the survey

This questionnaire survey was conducted to explore students' perspectives on what constitutes an ideal medical English textbook. By experiencing medical English learning through two different textbooks during the course, students were expected to develop clearer preferences and more informed opinions regarding medical English education.

2.2 Participants

At the end of the course in December 2024, we conducted the survey among first-year medical students who had used both textbooks. A total of 91 students participated.

2.3 Questionnaire

The questionnaire consisted of two sections. In the first section, students evaluated the two textbooks, EMS and BWC, on a five-point Likert scale across eight aspects: content intelligibility, sentence readability, vocabulary, exercises, visual support, unit structure, relevance to the national medical examination and clinical practice, and overall satisfaction with the textbook. The second section was open-ended, allowing students to freely describe the strengths and weaknesses of both EMS and BWC.

2.4 Analysis method

In the analysis, the first section was used to gain an overall impression of each textbook by comparing responses across the eight categories. Rather than conducting a direct comparison between the two textbooks, the focus

was on identifying the distinctive characteristics of each one. The open-ended responses from the second section were analyzed qualitatively to explore specific strengths and weaknesses, offering deeper insight into students' perceptions and supplementing the quantitative findings.

3. Results

3.1 Quantitative results: 5-point scale evaluation

The overall satisfaction rating for EMS was 4.30 out of 5, indicating a generally high level of student approval. The scores across the seven categories are presented in the following chart. With the exception of visual support (3.90), all categories scored above 4.00. The highest-rated aspect was *vocabulary*, which received a score of 4.24, followed by *content intelligibility*, which scored 4.20 (**Figure 1**).

In contrast, the overall satisfaction rating for BWC was 3.87, slightly lower than that of EMS. The detailed scores are shown in the chart below. Interestingly, unlike EMS, the highest-rated aspect for BWC was visual support, which scored 4.11. The lowest score was 3.91, given for relevance to the national medical examination and clinical practice. Another category that scored below 4.00 was vocabulary. In comparison with EMS, BWC received lower scores in all categories except visual support (**Figure 2**).

It is important to note that the purpose of this evaluation was not to directly compare the two textbooks, but rather to understand the distinct characteristics of each based on student feedback. Both EMS and BWC received generally high scores, indicating overall student satisfaction. However, the results revealed some notable differences. While the scores for BWC were slightly lower than those of EMS across most categories, *visual support* emerged as BWC's strongest feature and EMS's weakest. Conversely, *vocabulary* was the highest-rated category for EMS, but only the second-lowest for BWC.

3.2 Qualitative results: Free writing responses

In the free writing section of the questionnaire, we analyzed a total of 61 comments highlighting the strengths of EMS and 52 for BWC, along with 31 comments identifying weaknesses of EMS and 16 for BWC. Some responses were excluded from the analysis as they were not directly related to the textbooks themselves.

Upon reviewing the data, several recurring keywords and phrases emerged. These were categorized into common themes, and the frequency of each theme was counted. The results are as follows:

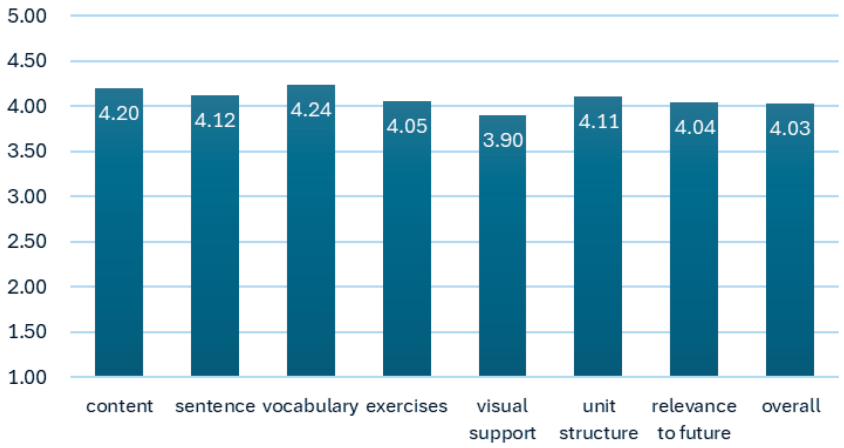


Figure 1. Scores across the seven categories of EMS

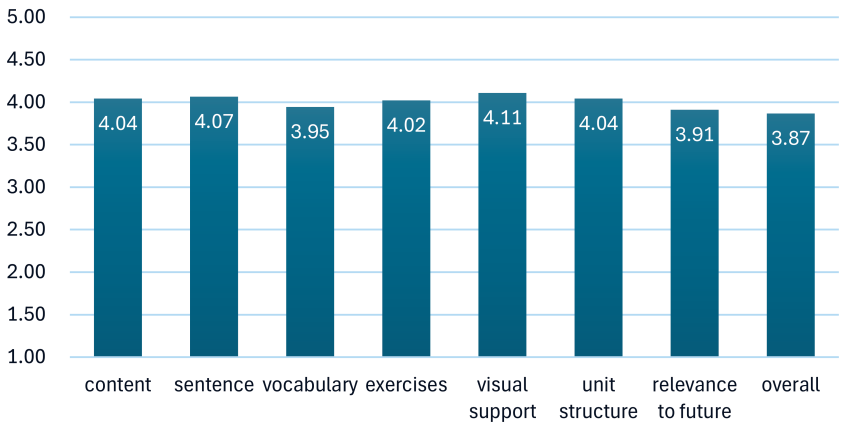


Figure 2. Scores across the seven categories of BWC

3.2.1 EMS – Strengths

The following keywords and phrases were extracted from students' comments on the strengths of EMS. The numbers in parentheses indicate how many times each was mentioned:

visual support (11), vocabulary (10), difficulty (8), anatomy (7), reading text (6), format/layout (5), content (4), interest (4), medical English (4), exercise (2), explanation (2), medicine (2), body structure (1), case (1), clinical scene (1), listening (1), medical knowledge (1), reading comprehension (1).

Overall, EMS was especially valued for its coverage of medical vocabulary, which many participants indicated was their first substantial exposure to such terms. Anatomy was also frequently mentioned (7 times), often highlighting the benefit of studying anatomy and English in parallel. Visual support, readability of the text, and clear formatting were consistently appreciated, with students noting that the level of difficulty felt appropriate and supported learning. Frequent references were made to its integration with anatomical content, which enhanced comprehension and reinforced lecture material (**Table 1**).

3.2.2 BWC – Strengths

The following keywords and phrases were extracted from students' comments on the strengths of BWC. The numbers in parentheses indicate how many times each was mentioned:

visual support (25), difficulty (11), listening (9), clinical scene (7), exercise (5), disease name (4), practical English (3), interest (2), communication (1), sentence (1), group work (1), vocabulary (1), pair work (1), familiarity (1).

Overall, BWC was most frequently praised for its strong visual elements (25 mentions) and listening exercises (9 mentions), which made the textbook more accessible and engaging. Students often linked these features to the theme of difficulty (11 mentions), indicating that the clear visuals and manageable listening tasks lowered the barrier to understanding. In addition, students valued its clinical scenarios and opportunities to learn practical English expressions, which were perceived as directly relevant to medical practice (**Table 2**).

3.2.3 EMS – Weaknesses

The following keywords and phrases were extracted from students' comments on the weaknesses of EMS. The numbers in parentheses indicate how many times each was mentioned:

error/outdated (12), visual support (7), difficulty (7), exercise (2), content (1), output (1), format (1), medical English (1), medicine (1), reading text (1), sentence (1), vocabulary/glossary (1), volume (1), Japanese explanation (1), explanation (1).

The most frequently cited concern was content accuracy and outdated information (12 mentions). Many students expressed sensitivity to the reliability of medical knowledge, emphasizing that inaccuracies in terminology or descriptions were problematic. Another recurring issue was insufficient or low-quality visual support (7 mentions), which students felt made the text harder to follow. The theme of difficulty (7 mentions) often overlapped with vocabulary load, use of medical English, or lack of supporting visuals, highlighting inconsistency in how students perceived the level of challenge (**Table 3**).

3.2.4 BWC – Weaknesses

The following keywords and phrases were extracted from students' comments on the weaknesses of BWC. The numbers in parentheses indicate how many times each was mentioned:

difficulty (7), listening (6), English ability/usefulness (3), Japanese explanation (1), anatomy (1), error (1), vocabulary (1), output (1), volume (1).

Table 1. Students' comments on the strengths of EMS

Keywords and Phrases	Students' Comments
Visual support (11 mentions)	• *The figures and English terms at the beginning were very easy to understand.* • *The anatomy diagrams at the beginning were easy to understand.*
Vocabulary (10 mentions)	• *I was able to learn medical English vocabulary.* • *Many medical English terms were included.*
Anatomy (7 mentions)	• *Reading the passages allowed me to review the knowledge I learned in anatomy.* • *It was easy to understand because it was related to the anatomy lectures.*
Reading text (6 mentions)	• *The sentences were simple and easy to understand.* • *The long passages were easy to understand and written in detail.*
Difficulty (8 mentions)	• *The content of the long passages was deep.* • *The passages were easy to read.*
Format/ layout (5 mentions)	• *The chapter structure was easy to understand.* • *The layout was easy to read.*

Table 2. Students' comments on the strengths of BWC

Keywords and Phrases	Students' Comments
Visual support (25 mentions)	• *Pictures helped me form an image.* • *It used diagrams, which were easy to understand, and allowed listening practice.*
Difficulty (11 mentions)	• *The passages were short and easy.* • *Many illustrations and diagrams made it very easy to understand.*
Listening (9 mentions)	• *The listening fill-in-the-blank tasks trained my ears.* • *The variety of conversations (listening exercises) was helpful.*
Clinical scene (7 mentions)	• *I could learn phrases used in medical practice.* • *The patient–doctor dialogues seemed useful for real practice.*

Table 3. Students' comments on the weaknesses of EMS

Keywords and Phrases	Students' Comments
Error/outdated (12 mentions)	• *Some terms were slightly different from what we learned in anatomy.* • *Some medical descriptions were outdated.*
Visual support (7 mentions)	• *More figures should have been included in the passages.* • *Some illustrations were hard to understand.*
Difficulty (7 mentions)	• *Without medical vocabulary knowledge, it was tough.* • *The vocabulary level should be lower.*

Table 4. Students' comments on the weaknesses of BWC

Keywords and Phrases	Students' Comments
Difficulty (7 mentions)	• *The listening tasks were far too simple.* • *It was a bit too easy; I might lose English skills.*
Listening (6 mentions)	• *The listening tasks were too easy.* • *The listening level didn't match.*
English ability /usefulness (3 mentions)	• *It might not really build English ability.* • *It may not be useful in the future.*

The primary criticism of BWC focused on its listening exercises, which were frequently described as too simple or poorly aligned with the learning content (6 mentions). The theme of difficulty (7 mentions) often appeared in connection with listening, reflecting students' sense that the material was not sufficiently challenging. Related to this, several students questioned the textbook's usefulness for improving English ability (3 mentions), expressing concern that because the tasks were relatively easy, they might not contribute to language development (**Table 4**).

4. Discussion

The present study examined student perceptions of two medical English textbooks currently used in the curriculum: *English for Medical Students* (EMS) and *Because We Care* (BWC). Both received generally favorable evaluations, but with distinct strengths and weaknesses that reflect their differing pedagogical orientations. The findings provide important insights into how students experience medical English learning materials, as well as implications for future curriculum design.

4.1 Key findings and interpretation

Quantitative results demonstrated a higher overall satisfaction rating for EMS (M = 4.30) compared to BWC (M = 3.87). EMS was particularly valued for its coverage of medical vocabulary and content intelligibility, while BWC was recognized for its visual support. These findings suggest that EMS is perceived as a more academically rigorous text, whereas BWC is seen as more accessible and engaging.

Qualitative analysis supported these trends. EMS was frequently praised for offering students their first substantial exposure to medical English terminology and for reinforcing anatomy lectures. Students highlighted the benefits of parallel study, noting that the integration of anatomical content enhanced comprehension. In contrast, BWC was appreciated for its clear illustrations, listening tasks, and clinical scenarios, which made the material approachable and relevant to communicative needs. Importantly, however, the two textbooks elicited different concerns: EMS was criticized for outdated or inaccurate medical content and insufficient visual support, while BWC was seen as too easy and unlikely to contribute to long-term English ability.

4.2 Comparison of textbook characteristics

The complementary nature of EMS and BWC was evident in both sets of data. EMS is academically oriented, emphasizing reading proficiency and vocabulary acquisition in alignment with subject matter content. This focus addresses the academic demands of medical education but may create challenges for students with limited English proficiency, particularly when medical terminology is presented without adequate glossing or explanation.

BWC, in contrast, is communicative in orientation. Its listening activities, clinical dialogues, and use of visual materials lower the barrier to participation, making it attractive to students who prefer interactive and accessible content. However, its perceived simplicity and lack of challenge indicate that students are also aware of the limitations of a textbook that prioritizes accessibility at the expense of linguistic rigor. Together, these findings reinforce the well-documented tension in ESP and CLIL contexts between subject content accuracy and linguistic accessibility (Dalton-Puffer, 2011; Dudley-Evans & St John, 1998).

4.3 Pedagogical implications

Several pedagogical implications emerge from this study. First, medical English textbooks must balance accuracy, relevance, and accessibility. The strong reaction to outdated or inaccurate content in EMS illustrates the importance of reliability in medical education materials. Even minor inconsistencies between textbook terminology and anatomy lectures undermined students' confidence. Updating and reviewing textbook content with medical specialists is therefore essential.

Second, the results point to the need for integrated materials that combine the academic rigor of EMS with the communicative accessibility of BWC. Students value both vocabulary acquisition and clinical application. A hybrid approach—embedding case-based dialogues and listening exercises into academically oriented chapters—may provide this balance. In addition, exercises that promote output (e.g., short writing tasks, oral practice) were rarely mentioned as strengths in either textbook and should be prioritized in future materials.

Third, the study highlights the role of teacher mediation. Part-time instructors or those without medical backgrounds may find BWC easier to use, while EMS requires more preparation and content knowledge. Developing comprehensive teacher's manuals and support resources would help ensure that instructors can deliver academically rigorous material effectively, regardless of their disciplinary expertise.

4.4 Limitations

This study has several limitations. The analysis was based on self-reported perceptions, which may reflect immediate impressions of difficulty rather than long-term learning outcomes. In addition, the evaluation was conducted within a single institutional context, and results may not generalize across medical schools with different curricula or student profiles.

4.5 Future directions

Beyond measuring outcomes, this study also points toward the next step of medical English textbook development. The analysis highlighted several important issues from the students' perspective, including the need for accurate medical content, clear visual support, and a balance between academic rigor and accessibility. These findings suggest that students are not only passive users but also valuable informants in shaping effective learning materials. Future studies should therefore continue to explore students' perspectives in more depth to identify the features that most enhance learning.

Materials development should focus on integrating academic content with clinical practice, ensuring accuracy while maintaining engagement. The incorporation of multimodal resources, updated medical cases, and scaffolded output tasks may address many of the concerns raised by students. By systematically reflecting student feedback in textbook design, future medical English resources can better support both the acquisition of specialized vocabulary and the development of communicative competence needed in clinical contexts.

4.6 Conclusion

In sum, both EMS and BWC received positive evaluations, but for different reasons. EMS was recognized for its academic depth and vocabulary support, while BWC was praised for its visual clarity and communicative orientation. Their weaknesses—accuracy issues in EMS and lack of challenge in BWC—highlight the need for materials that combine the strengths of both. By addressing these issues, future textbook development can better support medical students in acquiring the English proficiency required for both academic success and professional practice.

Acknowledgments

This study was supported by JSPS KAKENHI Grant Number JP24K04038.

References

1. Aono J and Considine D. 2006. *Introduction to medical English*［やさしい医学英語］. Igaku Shoin.
2. Dalton-Puffer C. 2011. Content-and-language integrated learning: From practice to principles? *Annual Review of Applied Linguistics* **31:** 182–204. https://doi.org/10.1017/S0267190511000092
3. Dudley-Evans T and St John MJ. 1998. *Developments in English for Specific Purposes: A multi-disciplinary approach*. Cambridge University Press.
4. Inoue M and Ihara T. 2020. *Because we care*. National Geographic Learning.
5. Kamiyama S, et al. (Year). *English for medical students*. Nan'un-do.
6. Kobayashi M. 2022. *Medical English for required medical English*. Kinpo Do.

Original article

Developing global competence in future healthcare professionals through a New Zealand study program (2009–2024)

Rie Sato,[1] John Telloyan,[2] and Jun Iwata[2]

[1] Department of Emergency and Critical Care Medicine, Faculty of Medicine, Shimane University
[2] Department of Medical English Education, Faculty of Medicine, Shimane University

Abstract

With the rapid advancement of internationalization, diverse educational programs are increasingly needed in medical and nursing education. However, overseas learning opportunities for medical and nursing students in Japan, particularly in the early years, remain limited. To address this, Shimane University has offered a comprehensive two-week study tour to New Zealand (NZ) for 1st- and 2nd-year medical and nursing students since 2009. This program aims to improve English communication skills, expand knowledge of NZ's medical system and practices, broaden perspectives on healthcare, and develop global awareness through immersive experiences. It also emphasizes the importance of English, motivating students to improve their English language skills. The structured schedule includes English lessons, lectures on NZ healthcare, visits to medical facilities, exchanges with local students and young doctors, homestays with NZ families, and weekend sightseeing trips. At the end of the tour, participants complete an evaluation questionnaire assessing their perceptions of the program. Analysis of past evaluations indicates that the tour not only improves English proficiency but also cultivates interpersonal skills through cross-cultural interactions and exposure to a different healthcare system. Early engagement with diverse cultures in medical and nursing education develops internationally minded professionals, encourages continued professional and language study, and broadens students' perspectives.

J Med Eng Educ (2026) 25 (1): 8-16

Keywords medical English education, overseas study tour, medical and nursing students, cross-cultural study, students' perceptions

1. Introduction

With the rapid advancement of internationalization, educational programs promoting global engagement have become increasingly available in Japan. With more Japanese students expressing interest in studying abroad, universities and colleges have expanded their overseas study offerings.[1, 2] Internationalization is particularly crucial in medical and nursing education, given the growing importance of English in these fields. English proficiency is essential for accessing medical knowledge, participating in international collaborations, and communicating with patients from diverse backgrounds.[3] These skills are vital for clinicians, researchers, and academics in today's global healthcare environment. As a result, medical and nursing schools have placed greater emphasis on English language education, leading to increased student interest in studying abroad or pursuing careers in English-speaking countries. However, study abroad opportunities remain limited, particularly for first- and second-year students. Despite the growing demand for international experiences, only 16.9% of medical schools offer overseas programs for first-year students, and 15.3% for second-year students.[4] A 2018 national survey on medical education reported that in 2017, just 128 first-year and 145 second-year students studied abroad, representing only 7.5% and 8.5%, respectively, of the 1,698 medical students who pursued international study that year.[5] This highlights a significant gap between student demand and

Corresponding author:

John Telloyan
English lecturer, Department of Medical English Education, Faculty of Medicine, Shimane University, 89-1 Enya-Cho, Izumo-shi, Shimane, 693-0021 Japan
TEL: 0853-20-2356 FAX: 0853-20-2340
E-mail: telloyan@med.shimane-u.ac.jp

Parts of this paper previously presented at *Advances in Language, Higher Education and Methodology 2025 International Conference* at Mongolian National University of Medical Sciences, Ulaanbaatar, Mongolia, on 20 March 2025.

program availability.

To address this issue, Shimane University launched a two-week study tour to New Zealand (NZ) in 2009, primarily for 1st- and 2nd-year medical and nursing students. The program enhances English skills, broadens students' perspectives as future healthcare professionals, and provides a comprehensive approach to medical English and internationalization. [6] Over the past 16 years (2009-2024), it has gained popularity among participants, with a total of 268 students having joined the program.

This paper aims to: (1) provide an overview of the study tour program developed by the Faculty of Medicine at Shimane University in collaboration with Waikato Institute of Technology (WINTEC) in NZ; (2) examine participants' perceptions of the program; and (3) discuss its impact on students' perspectives and subsequent medical and nursing studies.

2. Program overview

The overseas study program is a two-week initiative held in collaboration with WINTEC in Hamilton, NZ. The program is scheduled annually in March and accommodates up to 25 participants. It includes a range of activities designed to immerse students in an English-speaking environment while exposing them to NZ's unique healthcare and cultural practices.

2.1 Program objectives

This study tour program aims to achieve three primary objectives:

(1) To improve English communication skills necessary for a future career as a doctor or a nurse.
(2) To acquire knowledge of medical systems and practices in NZ and broaden perspectives concerning medicine and nursing.
(3) To enhance global awareness through overseas interactions and experiences, realize the importance of English, and motivate students to further their English language studies.

2.2 Participant recruitment

The number of participants in this study tour is set up to 25 for safety and quality control reasons. This program has been developed and targeted for 1st- and 2nd-year medical and nursing students at Shimane University. Applications are accepted on a first-come-first-served basis regardless of student English proficiency. While the program initially focused solely on nursing students in 2009, medical students have also been included since 2011. Recently, we usually have 2 to 5 nursing students, and more female students (75%) than male students (25%) among the 14-26 students every year.

2.3 Program development

This study tour program has been developed through collaboration between the Faculty of Medicine, Shimane University and WINTEC located in Hamilton, NZ. Both parties have agreed to a Memorandum of Understanding (MOU) and have enjoyed a strong, positive relationship. The International Centre of WINTEC has engaged in course development with both the Centre for Health and Social Practices and the Centre for Languages of WINTEC.

2.4 Program details

The duration of the study tour is approximately 20 days, including transportation, a two-week study program, and weekend sightseeing trips. This study tour has regularly been scheduled in March, the end of the Japanese school year. The program cost per participant amounts to approximately 450,000 - 600,000 JPY or $3,000-4,000 USD, covering tuition fees, accommodations (including homestays with three meals per day and hotels), travel insurance, weekend sightseeing trips, as well as all national and international travel expenses (air/train/bus tickets). The cost of the program has been increasing each year due to the impact of global inflation.

2.5 Supports for participants

Participants in this study tour are eligible for a scholarship, 50,000 JPY, offered by Shimane University. Additionally, a preparatory course consisting of ten 90-minute English lessons, taught by a native English language teacher, is provided to assist participants in preparing for their overseas experience. This course not only facilitates the development of English language skills necessary for the study tour but also imparts knowledge about NZ's culture, healthcare system, and safety measures. As part of the course requirements, participants are required to deliver presentations introducing Japanese culture, the Japanese healthcare system, and the Japanese medical/nursing educational system, which they later utilize to introduce Japanese culture to WINTEC staff and students.

2.6 Credit

This tour program is integrated into the medical/nursing curricula, and all participants receive school credit upon completion of the study tour.

2.7 Program schedule

The program's core components include:

(1) English for healthcare and a series of lectures on NZ healthcare: Morning English classes designed to familiarize students with medical terminology and afternoon classes for foundational knowledge about the NZ healthcare system
(2) Healthcare facility visits: Tours to various healthcare institutions, including general hospitals, clinics,

hospices, and birth centers, providing insights into NZ's holistic approach to care

(3) Cultural Experiences: Homestays with local families, interactions with Māori cultural events, and collaborative activities with WINTEC students and healthcare professionals

(4) Sightseeing Trips: Weekend trips to notable destinations such as Rotorua, Waitomo, and Lake Taupo, offering students a chance to explore NZ's natural and cultural heritage.

2.8 Number of participants and ratio of participants

In the past 16 school years (2009-2024), 268 students have participated, including 20 students who participated in online programs in the 2020 and 2021 school years. In 2019, the tour was canceled just before departure following a government advisory against overseas travel due to COVID-19. However, the program continued online in 2020 and 2021 (**Figure 1**).

As shown in **Figure 2**, the majority of participants were first-year medical students (M1), accounting for 54.9% of the total. Nursing students in their first to third years (N1–N3) made up approximately 31.4%, while second- and third-year medical students (M2 and M3) accounted for 13.8%.

2.9 Questionnaire Evaluation

On the final day of the tour each year, participants completed an evaluation questionnaire to assess their perceptions of the study tour. The evaluation structure consists of two main sections: A. Evaluation of the study tour program, including (1) English classes, (2) Medical and nursing studies, (3) Homestay, (4) Life in NZ, and (5) Support from Shimane University. B. Self-reflection on the study tour, including (1) English language skills, (2) Medical and nursing studies, (3) Cultural experiences and global perspectives, and (4) Self-management, cooperation, and leadership.

Each section contains four items for detailed evaluation, as outlined in **Table 1**.

The evaluation survey employs a Likert scale ranging from 1 (not satisfied at all / not improved or achieved at all) to 5 (very satisfied / significantly improved or achieved). The results from 2020 and 2021 were excluded as the programs were conducted online due to the COVID-19 pandemic and did not accurately reflect students' perceptions of the study tours held in NZ.

3. Results

3.1 Results of A. Evaluation of the Study Tour Program (satisfaction)

The results of the Evaluation of the Study Tour Program (satisfaction) are shown in **Table 2**.

(1) English classes:

The English classes section has consistently high ratings across all items, indicating overall satisfaction with the program.

a. contents: The average rating is 4.67 with a standard deviation of 0.17. The trend is decreasing over the years, suggesting potential areas for review and improvement.

b. facilities of the classroom: The average rating is 4.79 with a standard deviation of 0.12. The trend is stable with slight fluctuations, indicating well-maintained and suitable learning environments.

c. teacher(s): The average rating is 4.91 with a standard deviation of 0.07. The trend is stable with consistently high ratings, highlighting their effectiveness and positive impact on students.

d. support from Language Centre: The average rating is 4.78 with a standard deviation of 0.13. The trend is stable, indicating strong institutional support for

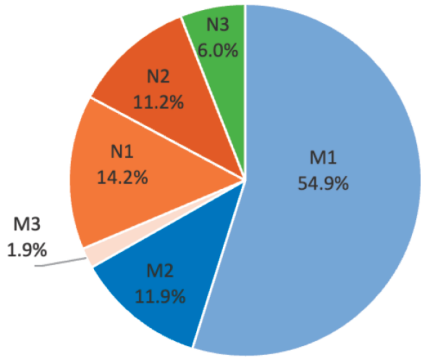


Figure 2. Ratio of participants

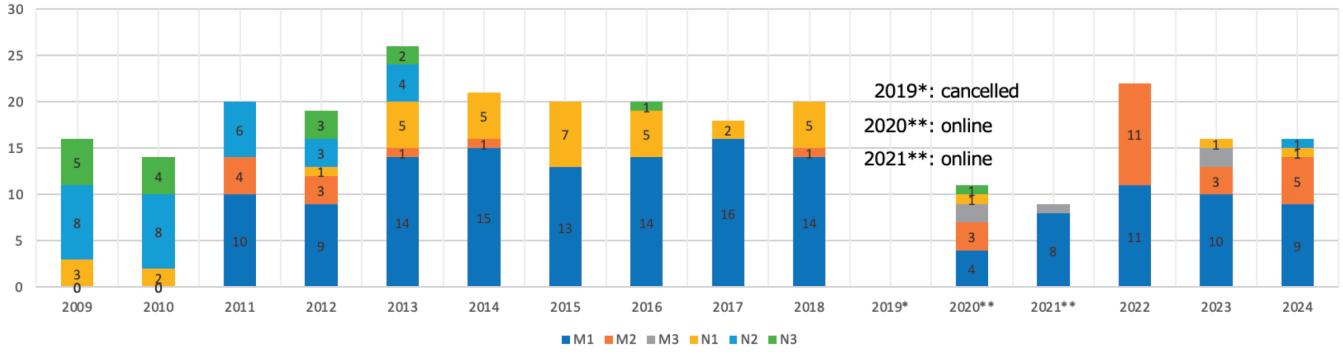


Figure 1. Number of participants from 2009-2024 school years

Table 1. Sections and items of the Evaluation questionnaire

A. Evaluation of study tour program (satisfaction)

"How satisfied were you with ...?"

Section	Item
(1) English classes	a. contents
	b. facilities of the classroom
	c. teacher(s)
	d. support from Language Centre
(2) Medical and nursing study	a. contents
	b. site visits
	c. teachers
	d. support from School of Health
(3) Homestay	a. house and facilities
	b. meals
	c. hospitality
	d. support from host family
(4) Life in NZ	a. life and safety
	b. sightseeing trips
	c. transportation
	d. culture and people
(5) Support from Shimane University	a. guidance seminars
	b. preparatory English lessons
	c. support from accompanying teachers
	d. guidebook

B. Self-reflection (improvement/achievement)

"How much do you think you could imporove/achieve...?"

Section	Item
(1) English language skills	a. reading
	b. writing
	c. listening
	d. speaking
(2) Medical and nursing studies	a. recognition of the importance of studying medical English
	b. interests in healthcare systems of other countries
	c. Recognition of differences in healthcare systems
	d. interactions with WINTEC staff and students
(3) Cultural understanding and global perspectives	a. cross-cultural understanding
	b. global viewpoints
	c. introduction of Japanese culture
	d. understanding NZ's culture
(4) Self-management, cooperation, and leadership	a. self-management
	b. group work
	c. leadership
	d. allowance management

Table 2. The results of the Evaluation of the Study Tour Program

Section	Item	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2022	2023	2024	avg.	SD
(1) English classes	a. contents	4.63	4.57	4.75	4.74	4.85	4.57	4.70	4.25	4.78	4.65	4.82	4.56	4.88	**4.67**	0.17
	b. facilities of the classroom	4.56	4.93	4.80	4.84	4.92	4.76	4.90	4.65	4.89	4.65	4.77	4.69	4.94	**4.79**	0.12
	c. teacher(s)	4.88	4.93	5.00	4.84	5.00	4.86	4.95	4.75	5.00	4.85	4.89	4.88	4.94	**4.91**	0.07
	d. support from Language Centre	4.69	4.79	4.90	4.74	4.88	4.67	4.85	4.55	4.89	4.55	4.90	4.81	4.94	**4.78**	0.13
(2) Medical and nursing study	a. contents	3.81	4.50	4.65	4.63	4.77	4.76	4.85	4.70	4.89	4.70	4.68	4.75	4.94	**4.66**	0.28
	b. site visits	3.38	5.00	4.60	4.68	4.92	4.62	4.75	4.85	4.72	4.80	4.59	4.50	4.86	**4.64**	0.41
	c. teachers	4.25	4.79	4.55	4.84	4.92	4.81	4.85	4.75	4.89	4.75	4.55	-	-	**4.72**	0.20
	d. support from School of Health	4.13	4.57	4.65	4.74	4.69	4.48	4.75	4.70	4.56	4.70	-	-	-	**4.60**	0.19
(3) Homestay	a. house and facilities	4.75	4.64	4.70	4.79	4.96	4.81	4.70	4.65	4.61	4.70	4.65	4.81	4.63	**4.72**	0.10
	b. meals	4.31	4.50	4.30	4.26	4.73	4.38	4.80	4.50	4.44	4.75	4.45	4.80	4.69	**4.53**	0.20
	c. hospitality	4.25	4.57	4.50	4.58	4.81	4.76	4.75	4.65	4.61	4.75	4.50	4.69	4.88	**4.64**	0.17
	d. support from host family	4.31	4.64	4.50	4.42	4.65	4.76	4.80	4.55	4.67	4.70	4.68	4.81	4.63	**4.63**	0.15
(4) Life in NZ	a. life and safety	4.63	4.64	4.65	4.58	4.92	4.76	4.95	4.75	4.50	4.00	4.50	4.88	4.75	**4.65**	0.25
	b. sightseeing trips	4.56	4.64	4.75	4.74	4.81	4.86	4.85	4.80	4.94	4.90	4.68	4.75	4.94	**4.79**	0.12
	c. transportation	4.25	4.50	4.70	4.47	4.73	4.57	4.45	4.40	4.39	4.15	4.23	4.44	4.44	**4.44**	0.17
	d. culture and people	4.75	4.79	4.90	4.89	4.92	4.95	4.90	4.75	4.89	4.70	4.77	5.00	4.81	**4.85**	0.09
(5) Support from Shimane University	a. guidance seminars	4.33	3.79	4.05	4.45	4.50	4.43	4.60	4.35	4.72	4.50	4.77	4.81	4.44	**4.44**	0.28
	b. preparatory English lessons	3.80	3.86	3.90	4.15	4.38	4.32	4.58	4.26	4.37	4.37	4.50	4.81	4.69	**4.31**	0.31
	c. support from accompanying teachers	4.63	4.43	4.80	4.60	4.88	4.71	4.80	4.90	5.00	4.85	4.82	4.88	4.81	**4.78**	0.15
	d. guidebook	4.63	4.50	4.85	4.60	4.85	4.84	4.89	4.79	4.74	4.79	4.50	4.81	4.56	**4.72**	0.14

language learning.

(2) Medical and nursing studies

The Medical and Nursing Study section shows a positive trend in ratings, particularly in the contents area, which has seen a steady increase over the years.

a. contents: The average rating is 4.66 with a standard deviation of 0.28. The trend is increasing over the years, indicating improvements or positive changes in the curriculum and course materials.
b. site visits: The average rating is 4.64 with a standard deviation of 0.41. The trend is stable with some fluctuations, indicating a generally positive but variable experience.
c. teachers: The average rating is 4.72 with a standard deviation of 0.20. The trend is stable, reflecting their competence and effectiveness.
d. support from School of Health: The average rating is 4.60 with a standard deviation of 0.19.

(3) Homestay

The Homestay section has high and stable ratings, indicating overall satisfaction with the homestay experience.

a. house and facilities: The average rating is 4.72 with a standard deviation of 0.10. The trend is increasing, suggesting improvements in the living conditions provided to students.
b. meals: The average rating is 4.53 with a standard deviation of 0.20. The trend is stable with some fluctuations, indicating variability in the dining experience.
c. hospitality: The average rating is 4.64 with a standard deviation of 0.17. The trend is stable, reflecting consistent and welcoming host families.
d. support from host family: The average rating is 4.63 with a standard deviation of 0.15. The trend is stable, indicating strong and reliable support for students during their homestay.

(4) Life in NZ

The Life in NZ section shows high and stable ratings, with an increasing trend in life and safety, indicating a positive and improving perception of safety and quality of life.

a. life and safety: The average rating is 4.65 with a standard deviation of 0.25. The trend is increasing,

indicating improvements or positive changes in the perception of safety and quality of life.

b. sightseeing trips: The average rating is 4.79 with a standard deviation of 0.12. The trend is stable, reflecting enjoyable and well-organized excursions.

c. transportation: The average rating is 4.44 with a standard deviation of 0.17. The trend is stable with some fluctuations, suggesting variability in the transportation experience.

d. culture and people: The average rating is 4.85 with a standard deviation of 0.09. The trend is stable with consistently high ratings, indicating a positive and enriching cultural experience for students.

(5) Support from Shimane University

The Support from Shimane University section shows generally high and stable ratings, with an increasing trend in guidance seminars, indicating improvements in the quality and effectiveness.

a. guidance seminars: The average rating is 4.44 with a standard deviation of 0.28. The trend is increasing, indicating improvements or positive changes in the quality and effectiveness.

b. preparatory English lessons: The average rating is 4.31 with a standard deviation of 0.31. The trend is stable, reflecting consistent quality in language preparation.

c. support from accompanying teachers: The average rating is 4.78 with a standard deviation of 0.15. The trend is stable with consistently high ratings, highlighting their crucial role in providing guidance and support to students.

d. guidebook: The average rating is 4.72 with a standard deviation of 0.14. The trend is stable, indicating its usefulness and reliability as a resource for students.

3.2 Results of B. Self-reflection (improvement/ achievement)

The results of the Self-reflection (improvement/ achievement) are shown in **Table 3**.

(1) English language skills

The English language skills section shows varying trends across different items, indicating areas of both improvement and potential review.

a. reading: The average rating is 3.66 with a standard deviation of 0.28. The trend is increasing, indicating improvements in reading skills over the years.

b. writing: The average rating is 3.56 with a standard deviation of 0.25. The trend is stable with some fluctuations, suggesting variability in writing skills.

c. listening: The average rating is 4.44 with a standard deviation of 0.26. The trend is increasing, indicating improvements in listening skills.

d. speaking: The average rating is 4.35 with a standard deviation of 0.30. The trend is stable with consistently high ratings, reflecting strong speaking skills.

(2) Medical and nursing studies

The Medical and Nursing Studies section shows a positive trend in ratings, particularly in the contents area, which has seen a steady increase over the years.

a. recognition of the importance of studying medical English: The average rating is 4.66 with a standard deviation of 0.24. The trend is increasing, indicating growing recognition of the importance of medical English.

b. interests in healthcare systems of other countries: The average rating is 4.66 with a standard deviation of 0.17. The trend is stable with consistently high ratings, reflecting strong interest in international healthcare systems.

c. recognition of differences in healthcare systems: The average rating is 4.56 with a standard deviation of 0.21. The trend is stable, indicating consistent recognition of healthcare system differences.

d. interactions with WINTEC staff and students: The average rating is 4.41 with a standard deviation of 0.22. The trend is stable with some fluctuations, suggesting variability in interactions.

(3) Cultural understanding and global perspectives

The Cultural Understanding and Global Perspectives section shows high and stable ratings, with an increasing trend in cross-cultural understanding and global viewpoints.

a. cross-cultural understanding: The average rating is 4.65 with a standard deviation of 0.22. The trend is increasing, indicating improvements in cross-cultural understanding.

Table 3. The results of self-reflection

Section	Item	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2022	2023	2024	avg.	SD
(1) English language skills	a. reading	3.31	3.64	3.20	3.53	3.77	3.62	3.85	3.60	3.61	3.80	3.45	4.31	3.88	**3.66**	0.28
	b. writing	3.06	3.93	3.20	3.53	3.65	3.71	3.80	3.55	3.61	3.45	3.36	3.88	3.56	**3.56**	0.25
	c. listening	3.94	4.07	4.25	4.47	4.38	4.81	4.60	4.50	4.56	4.30	4.36	4.81	4.69	**4.44**	0.26
	d. speaking	3.63	3.93	4.30	4.53	4.42	4.43	4.70	4.25	4.61	4.20	4.45	4.60	4.50	**4.35**	0.30
(2) Medical and nursing studies	a. recognition of the importance of studying medical English	4.19	4.19	4.60	4.68	4.73	4.52	4.85	4.80	4.78	4.65	4.73	4.88	4.94	**4.66**	0.24
	b. interests in healthcare systems of other countries	4.31	4.31	4.65	4.68	4.73	4.67	4.70	4.75	4.61	4.75	4.73	4.88	4.81	**4.66**	0.17
	c. Recognition of differences in healthcare systems	4.25	4.25	4.65	4.26	4.42	4.52	4.75	4.65	4.50	4.75	4.73	4.88	4.69	**4.56**	0.21
	d. interactions with WINTEC staff and students	4.25	4.25	4.60	4.47	4.50	4.48	4.40	4.35	4.39	3.80	4.55	4.63	4.63	**4.41**	0.22
(3) Cultural understanding and global perspectives	a. cross-cultural understanding	4.50	4.50	4.65	4.63	4.73	4.55	4.60	4.55	4.20	4.75	4.91	5.00	4.94	**4.65**	0.22
	b. global viewpoints	4.63	4.63	4.70	4.79	4.81	4.65	4.70	4.65	4.15	4.85	4.91	5.00	4.94	**4.72**	0.21
	c. introduction of Japanese culture	3.69	3.69	4.10	4.00	4.38	4.10	4.35	4.25	3.95	4.45	4.50	4.69	4.56	**4.21**	0.32
	d. understanding NZ's culture	4.38	4.38	4.45	4.58	4.54	4.40	4.75	4.50	4.15	4.75	4.91	4.75	4.81	**4.56**	0.22
(4) Self-management, cooperation, and leadership	a. self-management	4.44	4.44	4.45	4.47	4.50	4.57	4.40	4.65	4.56	4.55	4.38	4.31	4.19	**4.45**	0.12
	b. group work	4.13	4.13	4.55	4.47	4.35	4.63	4.58	4.58	4.32	4.84	-	-	-	**4.46**	0.23
	c. leadership	3.25	3.25	3.20	3.47	3.23	3.71	3.45	3.35	3.72	3.50	4.00	4.25	3.50	**3.53**	0.32
	d. allowance management	4.00	4.00	4.10	4.16	4.08	4.10	4.20	4.25	4.50	4.40	4.27	4.50	4.31	**4.22**	0.17

b. global viewpoints: The average rating is 4.72 with a standard deviation of 0.21. The trend is increasing, reflecting growing global perspectives.
c. introduction of Japanese culture: The average rating is 4.21 with a standard deviation of 0.32. The trend is stable with some fluctuations, suggesting variability in the introduction of Japanese culture.
d. understanding NZ's culture: The average rating is 4.56 with a standard deviation of 0.22. The trend is stable, indicating consistent understanding of NZ's culture.

(4) Self-management, cooperation, and leadership

The Self-management, Cooperation, and Leadership section shows generally high and stable ratings, with an increasing trend in leadership skills.

a. self-management: The average rating is 4.45 with a standard deviation of 0.12. The trend is stable, reflecting consistent self-management skills.
b. group work: The average rating is 4.46 with a standard deviation of 0.23. The trend is stable with some fluctuations, suggesting variability in group work skills.
c. leadership: The average rating is 3.53 with a standard deviation of 0.32. The trend is increasing, indicating improvements in leadership skills.
d. allowance management: The average rating is 4.22 with a standard deviation of 0.17. The trend is stable, reflecting consistent allowance management skills.

4. Discussion

4.1 Study tour program evaluation

(1) English Classes

The English classes received consistently high ratings across all evaluation items, indicating overall satisfaction with the program. However, a decreasing trend in the evaluation of course content suggests areas for improvement in materials and instructional methods. Classroom facilities and institutional support from the Language Centre were rated highly, reflecting a well-maintained learning environment. The consistently high ratings of teachers underscore their effectiveness and the positive impact on students' learning experiences.

(2) Medical and Nursing Studies

The evaluation of medical and nursing studies showed a positive trend, particularly in course content, indicating continuous improvements in the curriculum. Site visits and instructor support received stable ratings with some fluctuations, reflecting a generally positive but variable experience. The support provided by the School of Health remained consistent, ensuring students received the necessary assistance throughout their study tour.

(3) Homestay Experience

The homestay component of the study tour was highly rated and stable over the years, indicating overall satisfaction. Evaluations of house facilities showed an increasing trend, suggesting improved accommodations for students. While meal satisfaction fluctuated, hospitality and support from host families remained high, demonstrating a welcoming and enriching cultural experience for participants.

(4) Life in NZ

Participants reported a positive and improving perception of safety and quality of life in NZ, as reflected in increasing ratings in this category. Sightseeing trips were consistently rated highly, indicating well-organized and enjoyable excursions. Transportation received fluctuating ratings, suggesting some variability in public transit experiences. The consistently high ratings for cultural interactions indicate that students valued the opportunity to engage with NZ's people and traditions.

(5) Support from Shimane University

The support provided by Shimane University was rated highly and remained stable over the years. Guidance seminars showed an increasing trend in effectiveness, indicating improvements in preparation. Preparatory English lessons were rated consistently well, reflecting their utility in helping students develop language skills for the program. Support from accompanying teachers was highly valued, highlighting their crucial role in facilitating students' learning and adaptation.

4.2 Self-Reflection and Learning Outcomes

(1) English Language Skills

Students reported varying improvements in different language skills. Listening and speaking skills showed an increasing trend, indicating significant progress in these areas. In contrast, reading and writing skills remained stable with slight fluctuations, suggesting that these areas did not improve as much as shown in **Figure 3**. The overall ratings suggest that while the program effectively strengthened students' language abilities, additional focus on writing and reading practice could further enhance learning outcomes.

To statistically analyze the differences in average scores for listening, speaking, reading, and writing, an Analysis of Variance (ANOVA) test was conducted. It yielded an F-value of 15.67 and a p-value of 0.0001. Since the p-value is less than 0.05, we can conclude that there is a statistically significant difference between the average scores of listening, speaking, reading, and writing. These results indicate that while listening and speaking skills have shown significant improvement, reading and writing skills have not improved as much. Therefore, it is recommended to place additional emphasis on writing and reading practice to further enhance the overall language learning outcomes. Additionally, reporting effect sizes (e.g., η^2 or Cohen's f) would provide further insight into the magnitude of observed differences.

(2) Medical and Nursing Studies

The program contributed significantly to students'

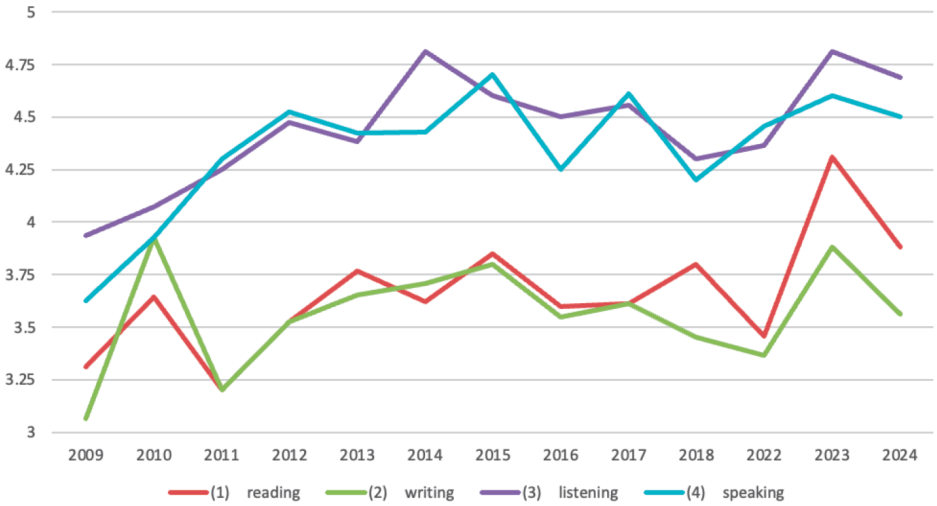


Figure 3. Trends in average scores of English language skills

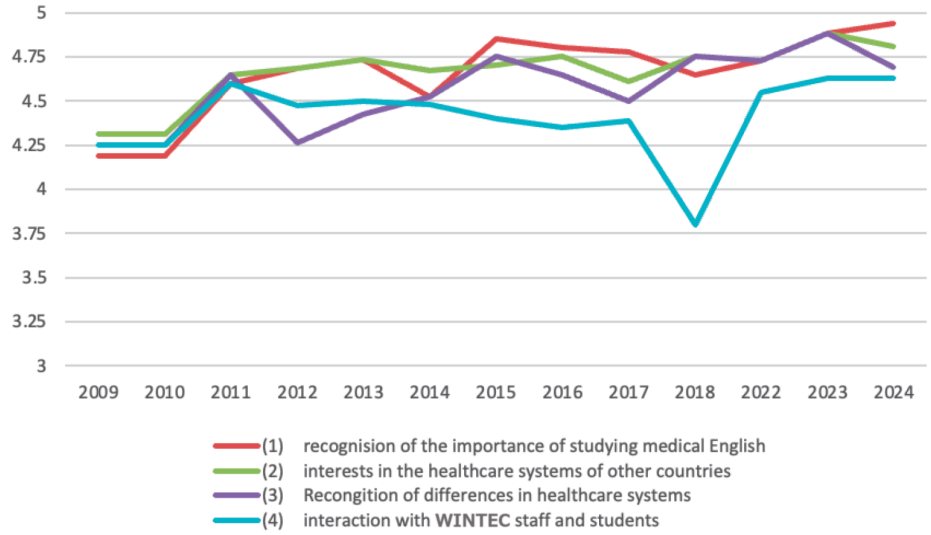


Figure 4. Trends in average scores of medical and nursing studies

recognition of the importance of studying medical English and increased their interest in international healthcare systems. The ability to recognize differences in healthcare practices between Japan and NZ was rated highly and remained stable as shown in **Figure 4**.

The ANOVA test yielded an F-value of 3.45 and a p-value of 0.02, indicating a statistically significant difference between the average scores of these items. These results suggest that while all items received high average scores, there are statistically significant differences in how students perceive the importance of studying medical English, their interests in healthcare systems of other countries, their recognition of differences in healthcare systems, and their interactions with WINTEC staff and students. Additionally, reporting effect sizes (e.g., η^2 or Cohen's f) would provide further insight into the magnitude of observed differences.

Additionally, it is noteworthy that the average scores for "interactions with WINTEC staff and students" are consistently lower compared to the other items across the years. This suggests that students may face more challenges or have less confidence in interacting with WINTEC staff and students. Possible reasons for this could include language barriers, cultural differences, or a lack of structured opportunities for meaningful interactions. Therefore, it may be beneficial to provide additional support and structured activities to facilitate better engagement and interaction between students and WINTEC staff.

(3) Cultural Understanding and Global Perspectives

Students reported increased cross-cultural understanding and global perspectives, with a steady upward trend. While students demonstrated an understanding of NZ's culture, their ability to introduce Japanese culture received slightly lower ratings, suggesting a need for more structured preparation in cultural exchange activities as shown in **Figure 5**. The ANOVA test yielded an F-value of 4.12 and a p-value of 0.01, indicating a statistically significant difference between the average scores of these items. Additionally, reporting effect sizes (e.g., η^2 or Cohen's f) would provide further insight into the magnitude of observed differences. Additionally, it is noteworthy that the average scores for "introduction of Japanese culture" are consistently lower compared to the other items across the

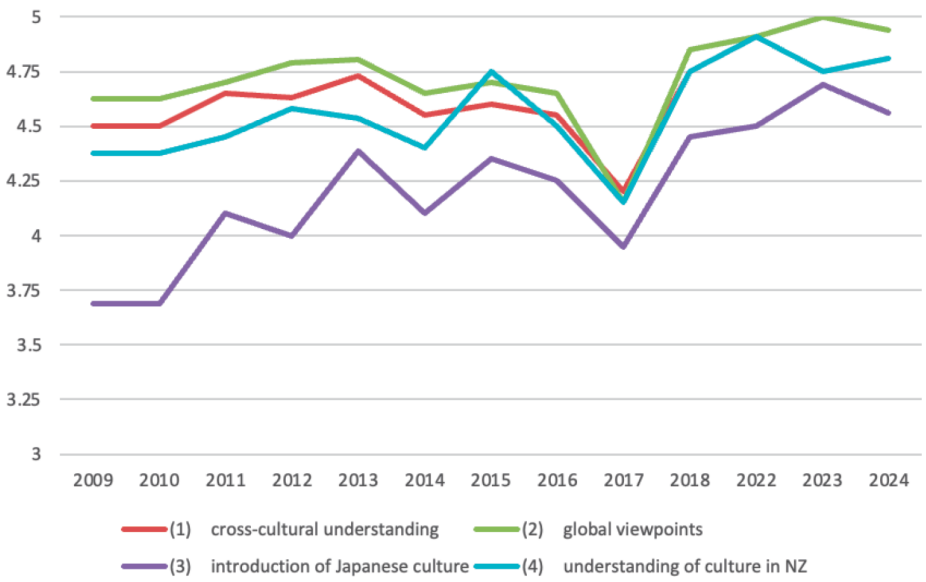


Figure 5. Trends in average scores of cultural understanding and perspectives

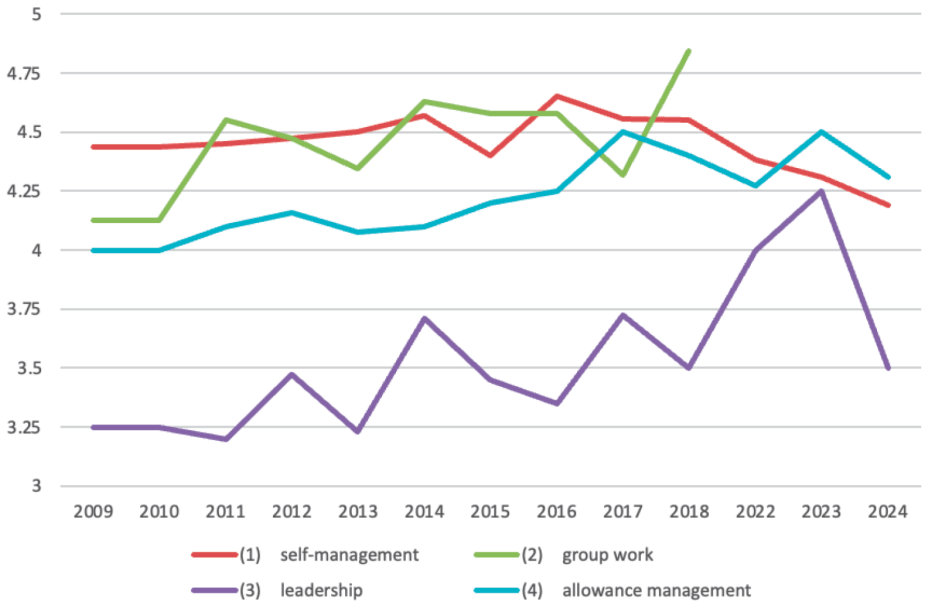


Figure 6. Trends in average scores of self-management and group work

years. This suggests that students may face more challenges or have less confidence in introducing Japanese culture compared to other aspects of cross-cultural understanding and global perspectives. Therefore, it may be beneficial to provide additional support and structured preparation in this area to help students improve their ability to share their cultural backgrounds effectively.

(4) Self-Management, Cooperation, and Leadership

As **Figure 6** shows, students rated their self-management and group work skills highly, with stable trends. However, leadership skills received relatively lower self-evaluation scores. This may be attributed to cultural factors influencing students' perceptions of leadership. Structured activities that encourage leadership development may help students gain confidence in this area.

4.3 Program strengths

4.3.1 Integrated Learning Approach

The combination of English classes, healthcare visits, and cultural activities created a well-rounded educational experience. Survey results indicated that students gained confidence in communicating in English and overcame their fear of making mistakes. Morning English lessons, focused on medical terminology and the NZ healthcare system, effectively prepared students for afternoon hospital visits. The structured learning environment helped students bridge the gap between their English proficiency and confidence in using the language in real-life scenarios.

4.3.2 Cross-Cultural Engagement

Exposure to NZ's multicultural society, including Māori traditions, interactions with local students, and insights

into holistic healthcare practices, broadened students' global perspectives. The opportunity to compare healthcare systems and interact with medical professionals in different settings encouraged students to reflect on Japan's medical system and their future roles as healthcare providers.

4.3.3 Collaborative Learning

The program's inclusion of both medical and nursing students fostered interdisciplinary learning and teamwork. The curriculum was intentionally designed to be shared by both groups, allowing students to complement each other's knowledge and skills. Survey results indicated that this interaction helped students develop a deeper appreciation for the collaborative nature of healthcare.

4.4 Recommendations for Improvement

4.4.1 Expanded Language Training

Survey results indicated that speaking and listening skills improved significantly, whereas reading and writing showed less progress. Given the short duration of the study tour and the emphasis on oral communication, it may be beneficial to incorporate structured reading and writing exercises into preparatory coursework to strengthen these skills.

4.4.2 Enhanced Cultural Communication Skills

Students expressed challenges in introducing Japanese culture to their NZ counterparts, likely due to a lack of confidence in their presentation skills. Strengthening preparatory sessions on cultural exchange and public speaking could improve students' ability to share their cultural backgrounds effectively.

4.4.3 Leadership Development

Students reported lower confidence in their leadership abilities despite having opportunities to lead group activities. This finding may reflect cultural tendencies toward modesty rather than an actual lack of leadership potential. Implementing structured leadership-building exercises could help students recognize and develop their leadership skills more effectively.

5. Conclusions

The overseas study program in New Zealand has proven to be a valuable initiative for first- and second-year medical and nursing students at Shimane University. Through language immersion, healthcare facility visits, and cross-cultural experiences, students gained a deeper understanding of international medical systems while enhancing their communication skills and global awareness. The findings suggest that early exposure to different healthcare environments can be instrumental in shaping students' perspectives on global healthcare practices.

Moving forward, implementing targeted improvements in language training, cultural exchange preparation, and leadership development will further enhance the impact of the program. Expanding collaboration with international institutions and incorporating digital learning tools could provide additional support for language and cultural preparation. Long-term follow-up studies could assess the sustained impact of the program on students' professional and academic trajectories, providing valuable insights for further program enhancements.

Ultimately, as globalization continues to influence the healthcare sector, initiatives like this study tour serve as a crucial foundation for developing internationally competent medical and nursing professionals. By continuing to refine and expand the program, Shimane University can further contribute to the cultivation of globally aware healthcare professionals who are prepared to navigate the complexities of international medical practice and cross-cultural patient care.

Acknowledgments

The authors are deeply indebted to everyone who supported students in this study program including their host families, students' families, and all individuals who supported the program. We are grateful to all staff members in the International Centre of WINTEC, the Centre for Health and Social Practices, the Centre for Languages in WINTEC, and the International Exchange Office, General Affairs Division, the Faculty of Medicine, Shimane University.

Declaration of interest

This research did not receive specific grants from public, commercial, or not-for-profit funding agencies.

References

1. Myers MW, Hiraizumi Y, Hashimoto M, and Miyazaki T. 2019. Study-abroad program for first-year students at Japanese comprehensive medical university. *J Med Eng Educ* **18**(2)**:** 32-36.
2. Bradford A. 2015. Changing trends in Japanese students studying abroad. *International Higher Education* **83:** 22-23.
3. Higa M, Davies W, and Fraser S. 2023. Investigating the English language needs of hospital medical team. *Hiroshima Studies in Language and Language Education* **26:** 17-30.
4. Noda C and Watanabe A. 2014. Future directions for curriculum development in English for medical purposes: A closer look at student needs. *J Med Eng Educ* **13**(3)**:** 65-74.
5. Association of Japan Medical Colleges. 2018. A white paper on the medical schools of Japan 2018. Japan: Kouhansha.
6. Telloyan J, Sato R, and Iwata J. 2022. Enhancing medical and nursing students' international studies through collaborative online international learning (COIL). *J Med Eng Educ* **21**(1): 49-52.

Original article

Enriching vocabulary learning through affixation: Classroom implications for medical English education

Haruko Shimazaki
Center for Language Education, Tsuru University

Abstract

Medical students are required to acquire a large volume of English terminology during their studies. Many of these terms are morphologically transparent, containing affixes such as *hyper-*, *hypo-*, *-itis*, and *-ectomy*. However, learners often rely on rote memorisation of isolated items and do not exploit morphological awareness for vocabulary expansion. This study examined the potential of affixation-focused activities to enrich learners' vocabulary and considered their applicability to English for Medical Purposes (EMP) instruction. Classroom activities targeting common prefixes and suffixes were trialled with intermediate-level EFL learners. Four activity types were introduced: (1) raising awareness of word families, (2) improving accuracy with negative prefixes, (3) addressing common meaning misconceptions (e.g., *inflammable*, *invaluable*), and (4) strengthening awareness of parts of speech. Learner responses and classroom interaction were examined qualitatively, with recurrent error patterns and learner reflections analysed. The findings indicate increased awareness of affix patterns, improved sensitivity to derivative forms, and reduced misconceptions about affixed vocabulary. Learner feedback suggested greater confidence in approaching unfamiliar words following explicit affixation-focused practice. While the intervention was exploratory and small-scale, the activities illustrate how affixation-based instruction may support vocabulary awareness relevant to English for Medical Purposes (EMP) instruction. Future research should extend this approach to medical and health science students and examine longer-term pedagogical effects.

J Med Eng Educ (2026) 25 (1): 17-22

Keywords affixation, English for Medical Purposes (EMP), learner strategies, morphology, vocabulary learning

1. Introduction

Medical and health science students in Japan are required to acquire a substantial amount of English terminology during their studies. Mastery of this terminology is essential not only for academic reading but also for professional practice, where accurate comprehension and production of English are needed for accessing international research and communicating with colleagues and patients.[1, 2] However, many students continue to rely on rote memorisation of isolated words, often using bilingual lists, with limited attention to morphological structure. Such approaches restrict their ability to expand vocabulary autonomously and to apply known items flexibly in new contexts.[3, 4]

Morphological awareness, particularly in terms of affixation, has been shown to facilitate systematic vocabulary expansion.[5-7] Affixation, the process of adding prefixes or suffixes to base words, is one of the most productive word-formation processes in English.[5] It provides learners with strategies for both decoding unfamiliar words and generating derivatives, and aligns with the Lexical Approach's emphasis on noticing and storing lexical patterns.[8] In English for Medical Purposes (EMP) instruction, many technical terms are morphologically transparent and thus potentially accessible through affix knowledge. For example, *hyperglycemia*, *hypotension*, *appendectomy*, and *nephrology* all contain affixes whose meanings can be systematically taught and applied.[6, 9] By recognising such patterns, learners can infer meanings, store related items in the form of word families, and use derivatives with greater accuracy.[7, 10] Recent studies have also emphasised the benefits of morphology-based instruction for medical vocabulary learning[11] and highlighted the importance of specifying essential academic and technical vocabulary for health science students.[12]

Nevertheless, research and classroom practice suggest that learners, even at intermediate and upper-intermediate levels, often underuse word-formation processes. They may rely on paraphrases rather than employing

Corresponding author:
Haruko Shimazaki
Center for Language Education, Tsuru University, 3-4-11-201 Tatsugaoka, Fujiyoshida-shi, Yamanashi, 403-0014 Japan
TEL: 090-6085-2311
E-mail: h-shimazaki@tsuru.ac.jp

appropriate derivatives (e.g., *did not understand* instead of *misunderstood*), select incorrect affixes (e.g., **unsignificant* instead of *insignificant*), or misinterpret affixed words (e.g., *inflammable* as "not able to burn"). Such tendencies have also been noted in tasks requiring transformation of base words, where learners frequently fail to produce the correct part of speech.[10]

The present study examined affixation-focused classroom activities designed to address these issues. Although the activities were implemented with general EFL learners rather than medical students, their design and observed effects are directly relevant to EMP. Many of the difficulties learners face with affixation in general English are also found in medical English, where precise understanding and use of affixes are crucial for comprehension and professional communication.

The objective of this paper is therefore twofold: first, to describe common learner problems regarding affixation and report on classroom activities trialled with intermediate EFL learners; and second, to discuss how such activities can be adapted and applied in medical English education. By highlighting both the observed learner responses and the potential transferability to EMP contexts, this paper seeks to contribute to ongoing efforts to improve vocabulary instruction for medical and health science students.

2. Methods

2.1 Participants

The classroom activities were implemented with 18 intermediate-level EFL learners (CEFR B1–B2) at a private language school in Sofia, Bulgaria. The group consisted of mixed-age adult learners (ages 17–62) with diverse professional and academic backgrounds. While the participants were not medical students, the tasks were designed with transferability to English for Medical Purposes (EMP) contexts in mind.

2.2 Materials

Instruction focused on common English prefixes and suffixes frequently encountered in both general and academic vocabulary. Tasks included word-family awareness, prefix-sorting, and pronunciation practice (see Appendix B for a sample worksheet and Appendix C for materials selected from *Formula B2 First* coursebooks and exam resources[14–16]).

1. Word family awareness (e.g., *happy → happiness, unhappy, happily*)
2. Negative prefix accuracy (*un-, in-, dis-, im-*)
3. Meaning misconceptions (e.g., *inflammable, invaluable*) supported with picture prompts
4. Parts of speech awareness through word formation tasks adapted from Cambridge English examinations

Although general vocabulary items were used in this trial, the same activity formats could be adapted to medical terms such as *hyperglycemia, appendectomy,* and *antibiotic*.

2.3 Procedure

The activities were conducted over two 60-minute lessons. The learners worked individually, in pairs, and in groups, depending on the activity. Each task followed a sequence of guided discovery, collaborative practice, and communicative application. For example, in the negative prefix task, the learners categorised adjectives under *un-, in-, dis-,* and *im-*, compared results in pairs, and then used the derived forms in short speaking tasks. Teacher monitoring and whole-class feedback addressed common errors and reinforced morphological patterns. The detailed sequence of classroom procedures is outlined in Appendix A (Lesson Plan Summary).

2.4 Data collection and analysis

Learner responses were collected through worksheets, classroom observation, and teacher field notes. Errors in derivative formation, prefix selection, and meaning interpretation were noted and categorised. Learner feedback was elicited informally at the end of each lesson regarding task difficulty and perceived usefulness. The analysis was descriptive, focusing on recurrent error patterns and learner reflections, rather than using inferential statistics, given the exploratory nature of the trial. Representative examples of learner tasks and worksheets are shown in Appendix C, and to situate these classroom activities within a principled pedagogical framework, the following section outlines the lesson design and its theoretical underpinnings.

No formal pre- or post-tests were administered; instead, the analysis focused on qualitative classroom data, including learner output, observed error patterns, and reflective feedback, in line with the exploratory teaching-focused nature of the intervention.

2.5 Lesson design and pedagogical rationale

This section summarises the lesson design and pedagogical rationale, clarifying how the materials were adapted and implemented. Affixation is one of the most productive word-formation processes in English and helps learners expand vocabulary systematically. Understanding how prefixes and base words combine allows learners to recognise more words in reading or listening and to produce new ones in speaking or writing.[9, 4, 12, 13] This is crucial for learners seeking greater accuracy and range in English use.

The target learners were Bulgarian adult intermediate students (CEFR B1–B2). Although they have studied vocabulary and grammar, they often rely on memorising individual words, which limits their autonomous development. The lesson addresses three key problems: underuse of word formation, incorrect prefix use (e.g., *umpossible*), and pronunciation issues (misplaced stress and

equal syllable weight).

The materials were selected from *Formula B2* (Pearson) and adapted for B1 learners. Exercises 1 and 2 were retained due to their accessibility and effective presentation of negative-prefix adjectives, while Exercises 3 and 4 were simplified by reducing word quantity and focusing on level-appropriate items. The lesson targets high-frequency negative prefixes (*un-*, *in-*, *dis-*, *ir-*, *im-*) that frequently appear in adjectives (e.g., *unhealthy, impatient, dissatisfied*). The topic and activities are suitable for a mixed-age, mixed-background group.

The lesson begins with a listening activity to introduce affixed adjectives in context, followed by explicit clarification of meaning, form, and pronunciation. Subsequent tasks progress from controlled to freer practice, reinforcing form accuracy and encouraging communicative use. Pronunciation is integrated at multiple stages, and collaborative pair work supports peer noticing and correction.

By the end of the lesson, the learners should be able to recognise, pronounce, and use adjectives with negative prefixes accurately and confidently. The sequence from noticing to production addresses the group's difficulties systematically, aligning with research suggesting that explicit morphology instruction enhances vocabulary learning and retention.[14–16]

3. Results

3.1 Accuracy in derivative formation

The classroom tasks revealed increased accuracy during guided practice, particularly after feedback and peer discussion. In the word family awareness tasks, once guided, most students successfully transformed base forms into nouns, adjectives, or adverbs (e.g., *analyze → analysis; happy → happiness; quick → quickly*). Common errors such as **successly* (for *successfully*) or **happyly* (for *happily*) were frequently observed at the beginning but became less prominent after feedback and peer correction.

3.2 Prefix selection

Frequent difficulties were observed in choosing the correct negative prefix. The learners initially produced forms such as **unsignificant* and **unloyal*; however, after sorting and categorisation practice, learners showed greater consistency in prefix selection during later stages of the activities, with most learners adopting the correct forms (*insignificant, disloyal*). Spelling issues such as **inpolite* instead of *impolite* were also reduced once rules (e.g., *in-* → *im-* before labials) were highlighted.

3.3 Misconceptions of meaning

Misinterpretation of affixed words was confirmed in items such as *inflammable* and *invaluable*. In the picture-choice activity, over half of the learners initially selected the incorrect meaning (*inflammable* as "not able to burn"). By the end of the activity, all the learners corrected their understanding and were able to generate their own example sentences demonstrating the correct meaning. These observations suggest successful conceptual clarification within the activity itself, though longer-term retention was not examined.

3.4 Awareness of parts of speech

In the Cambridge-style word formation tasks, the learners became more aware of the grammatical category required in word-formation tasks. Initially, many failed to transform items appropriately (e.g., using *product* instead of *producer*; *success* instead of *successful*). After explicit practice and collaborative discussion, most learners provided accurate derivatives, and several commented that they could better "see" the role of affixes in exam-style items.

3.5 Learner perceptions

Informal feedback suggested that the learners found the activities "challenging but engaging." Several participants reported that focusing on affixes helped them feel "more confident" when encountering unfamiliar vocabulary in texts. One noted that "it is easier to guess the meaning when I look at the beginning or the ending of a word." Another described the tasks as "useful, not only for exams but also for remembering words in groups, not one by one."

4. Discussion

The classroom activities reported in this study demonstrated that affixation-focused instruction can support learners' ability to form derivatives accurately, choose appropriate prefixes, and avoid common misconceptions. Even in a short series of lessons, the learners showed increased awareness of morphological patterns and greater confidence in approaching unfamiliar words. These findings are consistent with earlier research emphasising the role of morphological awareness in vocabulary development.[5-7, 11]

It is important to emphasise that the present findings do not imply sustained gains in productive morphological ability. Given the short duration of the intervention and its exploratory design, the observed effects should be interpreted as evidence of heightened awareness and task-based accuracy rather than durable learning outcomes.

Although the participants in this study were general EFL learners, the pedagogical implications are directly transferable to English for Medical Purposes (EMP). Medical terminology frequently incorporates affixes of Greek and Latin origin (*hyper-, hypo-, -itis, -ectomy, -ology*), and systematic instruction in these elements could help medical students decode complex terms more efficiently. For example, understanding the contrast between *hyperglycemia* and *hypo glycemia* requires only minimal base vocabulary knowledge

if the affixes are recognised. Similarly, awareness of suffixes such as *-ectomy* ("removal") or *-itis* ("inflammation") can significantly enhance comprehension of technical texts and support precise communication in clinical contexts.[1, 2, 11]

The results also suggest that affixation activities have the potential to foster learner autonomy. By learning to group words into families and predict meanings from form, students are less dependent on bilingual lists and can continue expanding their vocabulary beyond the classroom. This is particularly important for medical students, whose professional vocabulary needs are vast and continually expanding.[12]

However, several limitations must be noted. The study was exploratory, involved a small sample, and did not include medical or health science students. In addition, the participants were adult EFL learners in a European context, where exposure to English outside the classroom may be greater than in some other settings, which may have influenced their responsiveness to the activities. The activities were trialled only over two lessons, and no long-term retention data were collected. Therefore, while the findings indicate the promise of affixation-focused instruction, further research is needed to evaluate its sustained effects and direct applicability to EMP contexts.

Practical examples of these activities, including listening and speaking tasks, are provided in Appendices B and C to support classroom replication.

5. Conclusion

This exploratory classroom-based study illustrated how affixation-focused activities can support vocabulary awareness and task-level accuracy among intermediate EFL learners. Although the trial was conducted with non-medical students, the approach is highly relevant to medical English education, where affixation underpins much of the technical vocabulary. Integrating such activities into EMP curricula could help students decode, retain, and use medical terminology more effectively. Future studies should extend this line of research to medical and health science students, investigate long-term outcomes, and explore how affixation instruction can be combined with other strategies to meet the lexical demands of professional practice.

References

1. Ferguson G. 2013. *English for Medical Purposes.* In: Paltridge B, Starfield S, editors. The Handbook of English for Specific Purposes. Oxford: Wiley-Blackwell, pp. 243–261. doi: 10.1002/9781118339855.ch13
2. Coroban C. 2019. Teaching English to pharmacy students: Resources, importance, and applications. *Annals of Ovidius University of Constanța – Philology Series* **30**(1): 44–56.
3. Schmitt N. 2000. *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.
4. Nation ISP. 2013. *Learning Vocabulary in Another Language*. 2nd ed. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139858656
5. Bauer L and Nation P. 1993. Word families. *Int J Lexicogr* **6**(4): 253–279. doi: 10.1093/ijl/6.4.253
6. Schmitt N and Zimmerman CB. 2002. Derivative word forms: What do learners know? *TESOL Q* **36**(2): 145–171. doi: 10.2307/3588328
7. McCarthy M. 1990. *Vocabulary*. Oxford: Oxford University Press.
8. *Lewis M. 1993. The Lexical Approach: The State of ELT and a Way Forward.* Hove: Language Teaching Publications.
9. Crystal D. 1995. *The Cambridge Encyclopedia of the English Language*. 1st ed. Cambridge: Cambridge University Press.
10. Parrott M. 2000. *Grammar for English Language Teachers*. 2nd ed. Cambridge: Cambridge University Press.
11. Dang TNY and Webb S. 2014. The lexical profile of academic spoken English. *Engl Specif Purp* **33:** 66–76. https://doi.org/10.1016/j.esp.2013.08.001
12. Thornbury S. 2002. *How to Teach Vocabulary*. Harlow: Pearson Education.
13. Wang T and Zhang H. 2023. Examining the dimensionality of morphological knowledge and morphological awareness and their effects on second language vocabulary knowledge. *Front Psychol* **14:** 1207854. doi: 10.3389/fpsyg.2023.1207854
14. Pearson Education Limited. 2020. *Formula B2 First Coursebook*. Harlow: Pearson Education Limited.
15. Pearson Education Limited. 2020. *Formula B2 First Exam Trainer*. Harlow: Pearson Education Limited.
16. Pearson Education Limited. 2021. *Formula B2 First Teacher's Book and Teacher's Portal Access Code*. Harlow: Pearson Education Limited.

Appendices

Appendix A. Lesson Plan Summary (Affixation Lesson: Negative Prefixes)

Stage & Aims	Procedure	Outcomes	Resources & Materials	Interaction	Time (min)
1. Lead-in and Context Setting • Activate topic knowledge and prepare learners to use target language.	Show 2–3 images (e.g., products) to set context and elicit responses.Elicit brief reactions to activate schema.	Learners recall related vocabulary and anticipate topic.	Slides	WC	1–2
2. Listening (Gist) • Identify general topic and adjectives in context.	Play short audio; learners answer gist question and compare.	Learners identify topic and some adjectives.	Slides, audio	WC/SS/WC	3
3. Listening (Noticing Affixed Adjectives) • Raise awareness of form and meaning in context.	Learners listen again, noting adjectives with prefixes; compare in pairs.	Learners identify and highlight affixed adjectives.	Slides, Handout 1	WC/S/SS/WC	7–8
4. Presentation (Meaning, Form, Pronunciation) • Clarify meaning and pronunciation; introduce concept of word formation.	Present negative prefixes; clarify examples (e.g., *impatient, irregular*). Check understanding with CCQs; model pronunciation and stress.	Learners understand meaning and pronunciation of target words.	Slides, examples	WC	7–8
5. Controlled Practice 1 (Gap-fill) • Practise applying correct negative prefixes.	Learners add prefixes (un-, im-, in-, ir-, dis-) to adjectives and categorise them; pair check.	Learners form antonyms correctly; identify rules for prefix use.	Slides, Worksheet	S/SS/WC	7–8
6. Controlled Practice 2 (Pronunciation Focus) • Practise stress and weak forms.	Learners classify words by syllables, mark stress, and practise with audio.	Learners pronounce affixed words accurately.	Slides, Worksheet	S/SS/WC	7–8
7. Freer Speaking Practice • Encourage fluency and recall using target adjectives.	Learners discuss prompts using adjectives (e.g., *irrelevant, impatient*).	Learners personalise vocabulary and improve fluency.	Slides, card sets	SS/WC	10–15
8. Feedback • Consolidate learning and address errors.	Teacher elicits self-correction; summarises key points about affixation.	Learners reflect and internalise rules; improved awareness.	Slides	WC	1–2

Total Time: 43–50 minutes

Interaction codes: WC = whole class; SS = pair work; S = individual work

Appendix B. Sample Worksheet (Negative Prefixes)

Task 1. Listen and Complete the Sentences

Learners listen to a short audio and complete the sentences with the adjectives they hear.

1. People aren't always patient when they apply a fake-tanning product. They are often _____.
2. Some fake tans don't look natural. They look _____.
3. It didn't give a regular covering. It gave an _____ covering.
4. Using the product can be difficult if you aren't experienced. It can be hard if you're _____.
5. Many people weren't satisfied with the results. They were _____ with the results.

(Answers: *1. impatient, 2. unnatural, 3. irregular, 4. inexperienced, 5. dissatisfied*)

Task 2. Add These Words to the Correct Lines

Learners categorise adjectives under each prefix to notice spelling patterns and pronunciation shifts.

Prefix	Example Adjectives
un-	*affordable, appealing, interesting, reliable*
im-	*polite, possible, practical*
in-	*convenient, correct, sensitive*
ir-	*rational, regular, relevant, responsible*
dis-	*honest, satisfied, similar, organised*

After sorting, learners identify spelling rules (e.g., *im-* before labials, *ir-* before r-words) and create one original sentence using a word from each prefix group.

Task 3. Pronunciation & Stress Awareness

Learners classify words by syllable count, then mark the stressed syllable and practise aloud. They compare base and derived forms, noting that prefixes are usually unstressed.

Three Syllables	Four Syllables	Five Syllables
dishonest, impatient, impolite, incorrect, unhealthy, unnatural	*dissatisfied, irrelevant, uninteresting*	*inexperienced, irresponsible*

Extension: Learners underline the stressed syllable and practise choral and individual repetition.

Examples

dishonest / dɪs'ɒnɪst/ → prefix weak

uninteresting / ˌʌn'ɪntrəstɪŋ/ → stress on second syllable

Appendix C. Supplementary Classroom Activities (Freer Practice)

1. Speaking Prompts – "Guess My Word!"

Game-based fluency activity where learners describe an adjective without saying it.

Example

"You say this when someone doesn't tell the truth" (Answer: "*dishonest*")

Learners take turns. Encourages lexical recall, spontaneous use, and peer correction.

2. Word Family Challenge

Learners choose a base word and collaboratively create as many derivatives as possible.

Example

"Satisfy" (Answers: *satisfaction, dissatisfied, satisfying, satisfactorily*)

They explain differences in meaning and form.

3. Matching Cards for Word Formation Practice

Pairs match adjectives with their base forms (e.g., *patient → impatient; honest → dishonest*), reinforcing morphological awareness and productive accuracy.

4. Reflection Prompt

Learners write a short paragraph about which prefixes they find most confusing and why, discussing pronunciation or spelling challenges with partners.

Original article

Refining the Japan Functional History Taking Assessment (JFHTA): A pilot study to optimize interrater reliability using PDCA cycles

Eric H. Jego,[1, 4] Kazunori Kawasaki,[2] Wipakorn Chayopong Kawasaki,[3] and Tetsuya Taniguchi[4]

[1] Medical Education Center, Nihon University School of Medicine
[2] General Studies, Nihon University College of Engineering
[3] Language Institute, Thammasat University
[4] Division of Liberal Arts, Nihon University School of Medicine

Abstract

Background: As the number of international patients in Japan increases, there is a growing need to assess the English clinical communication skills of Japanese medical students. However, existing assessment tools often conflate linguistic proficiency with clinical competence or rely on subjective judgments that fail to achieve interrater reliability.

Objective: This study aimed to refine the Japan Functional History Taking Assessment (JFHTA) and evaluate the feasibility of improving its interrater reliability through an iterative design process.

Methods: The study utilized a Plan-Do-Check-Act (PDCA) framework across four development phases (v8.3 to v9.6). Three experienced and diverse English educators evaluated 10 videos of medical history-taking performances across four PDCA cycles (total: 30 evaluation events). Reliability was measured using the Intraclass Correlation Coefficient (ICC) with 95% Confidence Intervals (CI).

Results: The Pilot phase (v8.3) revealed Poor reliability (ICC < 0) due to structural confusion between interactional and investigative assessment criteria. Through iterative refinements such as the structural prioritization of interactional patient-centeredness, reliability improved. The final version (v9.6) achieved improved reliability (ICC = 0.92). However, CI remained wide (95% CI: 0.47 - 1.00), reflecting the limited sample size.

Conclusion: While statistical precision was limited by sample size, the PDCA framework successfully identified and addressed sources of rater disagreement. The study demonstrated that engaging diverse raters in continuous improvement cycles can resolve interpretative discrepancies. This provides preliminary evidence that JFHTA v9.6 has the potential to reliably assess medical history taking in novice pre-clinical medical students.

J Med Eng Educ (2026) 25 (1): 23-29

Keywords communication, interrater reliability, medical history taking, quality improvement, undergraduate medical education

1. Introduction

1.1 English for Medical Purposes (EMP) education in Japan

With the rising number of international residents and visitors in Japan, the need for Japanese medical doctors to communicate in English is greater than ever.[1] Factors including globalization, fluctuating economic conditions, and record-breaking inbound tourism are contributing to a growing number of international patients requiring English language support to access Japanese healthcare.[2,3]

In recent years, Japanese medical schools have been investing more resources and effort toward improving students' clinical communication skills.[4] The current state of English education in Japanese medical schools is often focused on reading and terminology, and less on functional clinical skills. However, the number of EMP programs is increasing, including a growing interest in clinical communication in English. Despite a robust OSCE system

Corresponding author:

Eric Hajime Jego
PhD, PGCertMedEd, Director of English for Medical Purposes Education, Medical Education Center/Division of Liberal Arts, Nihon University School of Medicine, 30-1 Oyaguchi Kamicho, Itabashi-ku, Tokyo, 173-8610 Japan
TEL: 03-3972-8111 (ext. 2312)
E-mail: jego.erichajime@nihon-u.ac.jp

implemented uniformly nationwide in Japan, no nationally standardized protocols or guidelines exist to assess oral communication skills in EMP.[1]

1.2 Assessment gap

In order to appropriately measure oral communication skills, there are several factors which need to be taken into careful consideration, including the target students, their existing knowledge, and the current learning environment. The student videos examined in this study were those of pre-clinical novice Japanese medical students with limited medical knowledge, playing the role of a doctor with a simulated patient. As part of their compulsory coursework, they have an English for Medical Purposes course focused on developing medical history-taking skills from a patient-centered perspective. The course is active and output-based with programmatic assessment implemented to support outcome-based education as mandated by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) through the Model Core Curriculum for medical education in Japan.[5] Given the lack of medical training and expertise, the course does not prioritize clinical knowledge and skills. Rather, it focuses on developing patient-centered communication (Interactional criteria) and collecting medical information from patients (Investigative criteria), so a suitable assessment tool is required.

For this cohort, existing assessment tools are excessively linguistic (e.g., TOEIC, TOEFL) and focus on English mechanics such as grammar, vocabulary, and comprehension, where topics are rarely relevant to the study of medicine. Alternatively, in medicine, there are many existing evaluation options (e.g., OSCE, mini-CEX, former USMLE Step 2 CS), but they were all designed to measure clinical skills and assume prior clinical skills training, which is also inappropriate for this cohort.

1.3 Development of JFHTA

In response to this need, the Japan Functional History Taking Assessment (JFHTA) was developed as part of a doctoral dissertation in 2022.[6] JFHTA was a pedagogically motivated project built through a series of cyclical iterations to improve its design and implementation. Each iteration of JFHTA was trialed, data was analyzed, improvements were made to create subsequent versions from version 1 (v1.0) concluding with version 8 (v8.0) with the informal input of trusted colleagues along the way.

JFHTA was designed to fulfill the need to appropriately assess the oral communication skills of novice Japanese medical students within the context of medical history taking. JFHTA objectively measures student utterances loosely categorized in accordance with the Roter Interaction Analysis System (RIAS)[7] and principles of Interactional Competence.[8] It was originally developed through an intensive process which was scrutinized by leading applied linguistics experts over a period of 10 years involving several hundred individual assessment events. To date, no studies exist to establish whether JFHTA can function reliably when multiple evaluators assess the same student.

1.4 Study objective

This study sought to answer the following question: Can a PDCA-style iterative development process involving raters from different backgrounds result in a more reliable assessment tool? This study shows the development of JFHTA from v8.3 to v9.6 including data analysis obtained through an extensive process of improvement cycles to assess potential improvements in interrater reliability with a diverse team of 3 experts in English education.

2. Methods

2.1 Study design

This study utilized a fully crossed, quantitative research design to assess the interrater reliability (IRR) of the Japan Functional History Taking Assessment (JFHTA) across multiple iterative development cycles. The study was conducted within the context of a required English for Medical Purposes (EMP) course at a private medical university in Tokyo, Japan.

The research was structured around the principles of Quality Improvement (QI), specifically utilizing Plan-Do-Check-Act (PDCA) cycles. Unlike a static reliability study, which measures a tool at a single point in time, this study was dynamic; the assessment tool itself was the primary variable, undergoing modifications between data collection phases based on rater feedback and statistical performance. The primary objective was to calibrate the tool until it achieved acceptable reliability coefficients among a linguistically and professionally diverse group of raters.

2.2 JFHTA: instrument structure

JFHTA measures history-taking performance across five parts: A, B, C, D, and E. Preliminary analysis showed that Parts C (Undesirable Behaviors), D (Overall English), and E (Overall Impression) achieved acceptable interrater consensus early in the study; thus, they were excluded from detailed statistical analysis. Reliability analysis focused on the two sub-scales responsible for the majority of rater variance, A and B, as described below for the current version of JFHTA v9, as well as the Total Score:

- Part A Interactional (Patient Centeredness and Communication): This part (formerly known as Non-Medical) assesses expressions intended to build patient rapport and to facilitate communication. It focuses on patient-centered behaviors, including expressing empathy, intention, and understanding, as well as asking for permission and patient opinion and input. It also encompasses interactional aspects

of conversation, such as overcoming communication difficulties, clarifying information, and active listening skills.

- Part B Investigative (Information gathering): This part (formerly known as Medical) assesses the medical information gathering function of history taking (e.g., eliciting the chief concern, asking about the history of present illness). It focuses on what the student asks and rewards them for asking follow-up questions which include context provided by the patient.
- Total Score: The aggregate performance metric combining all five parts of JFHTA.

The total score is interpreted using an accompanying JFHTA rubric defining bands of performance into letter grades, S, A, B, C, D, and D-, and an additional grade of Incomplete for those who did not attempt the roleplay activity. It is notable that the earlier version of JFHTA (v8) had a different sub-structure (Parts A1/A2, B, C, D, and E), which was largely redesigned to make v9.

2.3 Participants and materials

2.3.1 Video selection

A total of 10 video recordings were utilized for the analysis (n = 10): nine from first-year Japanese medical students and one expert calibration video (a British physician) used in the pilot phase. Student recordings were collected from final EMP course evaluations and anonymized according to institutional ethical guidelines featuring learners roleplaying with the lead researcher acting as the patient.

A purposeful sampling strategy was employed to select student videos that represented borderline cases as defined by the JFHTA rubric. Selection was performed by Rater 1 based on the original course evaluations. Rather than selecting clearly superior or clearly inferior performances, which are often less challenging to grade consistently, videos were specifically chosen because they straddled the boundaries between grade bands (e.g., performances falling between S/A, A/B, or B/C grades). While this introduces selection bias based on Rater 1's prior assessment, this selection criterion was intended to rigorously stress-test JFHTA's sensitivity and its ability to distinguish between proficiency levels in ambiguous scenarios.

2.3.2 The raters

The assessment team consisted of three fixed raters (k = 3) who evaluated different sets of videos in each phase (Phase 1: n = 2; Phase 2: n = 2; Phase 3: n = 2; Phase 4: n = 4). While the limited sample size in the early phases (n = 2) introduces statistical volatility, these phases served primarily as diagnostic pilots to identify structural failures in the rubric. This fixed-rater design was chosen to isolate the rubric text variable; any changes in reliability could be attributed to improvements in the tool rather than changes in the rater pool. The raters represented a diverse range of linguistic and professional backgrounds:

- Rater 1: Developer of JFHTA and a Canadian medical education researcher and educator based in Japan. Rater 1 was the course instructor and was not blinded to the students' identities.
- Rater 2: A Japanese expert in English for Specific Purposes (ESP) education within the field of Engineering. Rater 2 was blinded to student identities and grades.
- Rater 3: A Thai national and university lecturer in English as a Foreign Language (EFL). Rater 3 was also blinded to student identities and grades.

2.4 Procedure: PDCA cycles

The data collection was organized into four distinct phases spanning approximately a two-year period, corresponding to versions of the JFHTA: Phase 1 (Pilot/v8.3), Phase 2 (v9.3), Phase 3 (v9.5), and Phase 4 (Final/v9.6). Each phase was marked by a PDCA cycle in accordance with best practices in medical education in Japan.[9]

- Plan: The team identified specific areas of the JFHTA scoresheet that caused scoring discrepancies in previous trials (e.g., vague descriptors in the Interactional section; misunderstood descriptors in the Investigative section, etc.).
- Do: The raters watched the designated videos and assigned scores using the current version of JFHTA. Raters did not communicate during the viewing process.
- Check: Immediately following the scoring, the three raters compared and discussed results in person. This phase was qualitative and discursive. When score discrepancies arose, the team re-watched relevant portions of the video together. The goal was not to force consensus on the score, but to identify why the tool allowed for such divergent interpretations.
- Act: Based on the analysis of the discrepancies, the text of the JFHTA was revised. Ambiguous words were replaced, sections were redesigned, and categories were streamlined. The revised version was then prepared for the next cycle.

2.5 Statistical analysis

2.5.1 Intraclass Correlation Coefficient (ICC)

To assess interrater reliability, we utilized the Intraclass Correlation Coefficient (ICC). We selected the ICC(2,1) model (Two-Way Random, Single Measures, Absolute Agreement) based on the guidelines by Shrout and Fleiss[10] and McGraw and Wong.[11]

- A Two-Way Random model was chosen to account for variance in both the video difficulty and the raters themselves, assuming both are drawn from larger populations.

- Single Measures was selected over Average Measures to simulate the ecological validity of a classroom setting where students are typically graded by a single examiner rather than a panel.
- Absolute Agreement was prioritized over Consistency to ensure that raters assigned identical scores, important for standardized pass/fail decisions.

2.5.2 Calculation and interpretation

Data management and analysis were initially performed using Python (v3.10) with the pandas and statsmodels libraries, then verified by a professional statistician (T.T., co-author). The ICC was calculated based on ANOVA Mean Squares. In addition to point estimates, 95% Confidence Intervals (CIs) were calculated to assess the precision of the reliability estimates given the small sample size. Interpretation of ICC values followed the standard benchmarks suggested by Koo and Li:[12]

- < 0.50: Poor reliability
- 0.50 – 0.75: Moderate reliability
- 0.75 – 0.90: Good reliability
- > 0.90: Excellent reliability

The analysis tracked ICC values for the Interactional, Investigative, and Total Scores across all four phases to evaluate the progressive improvement in reliability resulting from the iterative PDCA refinements.

3. Results

3.1 Overview of analysis

The reliability analysis was conducted on a total of 10 video performances (n = 10) evaluated by three fixed raters (k = 3), resulting in 30 individual assessment events. The analysis tracked the Intraclass Correlation Coefficient (ICC) across four distinct phases of the JFHTA development (v8.3 to v9.6).

The results demonstrate a substantial increase in interrater reliability for the Total Score across the four phases. As PDCA cycles progressed, JFHTA evolved from Poor reliability in the Pilot phase (ICC < 0) to Excellent reliability in the Final phase (ICC = 0.92).

3.2 Phase 1: Pilot (v8.3)

The initial analysis of the Pilot phase, which included one calibration video with a working clinician and one student video, revealed significant structural flaws in the assessment tool. Reliability was poor across all domains, with ICC values dropping below zero (ICC < 0 for Interactional, Investigative, and Total Scores). These negative values indicated systematic disagreement among raters; specifically, the variance due to rater differences exceeded the variance between the actual performances. Rater feedback during the "Check" phase suggested that definitions of "patient-centeredness" were interpreted differently based on the raters' cultural backgrounds and expertise (Western vs. Japanese vs. Thai). Additionally, the Interactional criteria appeared in two different sections in v8.3, which was unnecessarily confusing for both Rater 2 and Rater 3. Furthermore, Investigative utterances (i.e., those intended to gather new medical information) were often misinterpreted by the same two raters. Due to all of these deficiencies, a major redesign of JFHTA was required.

3.3 Phase 2 (v9.3) and Phase 3 (v9.5)

Following the rebuild, Phase 2 (v9.3) showed the first signs of rater agreement. The Total Score improved to Moderate reliability (ICC = 0.53). The Investigative Score also reached Moderate reliability (ICC = 0.54). However, the Interactional Score remained problematic (ICC = -0.10), suggesting that while raters agreed more on Investigative element scoring, they continued to disagree on Interactional element scoring.

Phase 3 (v9.5) marked a continuing evolution of the calibration process. The Total Score demonstrated Excellent reliability (ICC = 0.95), driven largely by the Investigative Score (ICC = 0.93). Conversely, the Interactional Score remained at 0.00. However, the wide confidence interval (-0.31 to 1.00) indicates that this low value was attributable to the small sample size (n = 2) and lack of variability in performance, causing a calculation artifact, rather than indicating rater disagreement as the scores showed very close agreement.

3.4 Phase 4: Final (v9.6)

The final iteration (v9.6) incorporated specific behavioral indicators for Part A Interactional, which were clearer and simpler. In this final phase, reliability was achieved across all domains. The Interactional Score achieved Excellent reliability (ICC = 0.96). Notably, despite the small sample size, the lower bound of the 95% Confidence Interval for the Interactional Score remained high (0.82), suggesting robust agreement on communicative behaviors. The Investigative Score maintained a Good reliability level (ICC = 0.77). Consequently, the Total Score for the final version of JFHTA achieved an ICC of 0.92. However, as shown in **Table 1**, the confidence intervals for the Total Score remained wide (0.47 to 1.00), indicating that while the Total Score ICC reliability is Excellent, statistical precision is limited by the sample size.

4. Discussion

4.1 The rebuild: from structural confusion to functional design

The reliability coefficients observed in this study suggest that the initial failure in the Pilot phase was less about rater error and more about instrument design. The extremely negative ICC values in this early phase indicated that the

Table 1. Intraclass Correlation Coefficients, ICC(2,1) and 95% Confidence Intervals (CI) across JFHTA v8.3 to v9.6.

JFHTA Version *(number of videos)*	Interactional ICC *(95% CI)*	Investigative ICC *(95% CI)*	**Total Score ICC** *(95% CI)*	**Total Score Reliability Interpretation**
v8.3 ***(n = 2)***	-0.57 *(-0.68, 0.79)*	-1.06 *(-1.19, 0.69)*	**-0.88** *(-0.95, -0.34)*	**Poor**
v9.3 ***(n = 2)***	-0.10 *(—, 0.10)*	0.54 *(-0.01, 0.99)*	**0.53** *(0.01, 0.99)*	**Moderate**
v9.5 ***(n = 2)***	0.00 *(-0.31, 1.00)*	0.93 *(0.24, 1.00)*	**0.95** *(0.36, 1.00)*	**Excellent**
v9.6 ***(n = 4)***	0.96 *(0.82, 1.00)*	0.77 *(0.25, 0.98)*	**0.92** *(0.47, 1.00)*	**Excellent**

Note. ICC = intraclass correlation coefficient. Reliability interpretation based on Koo and Li guidelines: < 0.50 = Poor; 0.50 to 0.75 = Moderate; 0.75 to 0.90 = Good; > 0.90 = Excellent.12
Dash (—) indicates the confidence interval could not be calculated due to zero variance or small sample size.

tool itself was generating systematic error. Version 8.3 of the JFHTA scoresheet was originally organized into two parts: when the doctor is speaking (Part A) and when the doctor is listening (Part B). In version 8.3, raters were required to categorize short active listening cues in Part B, while longer patient-centered utterances were scored in a separate sub-section (Part A2). Raters found it mentally challenging to navigate between these different sections while tracking the live performance. Furthermore, this layout failed to clearly distinguish between gathering new information and confirming old information. Raters frequently struggled to categorize questions that served a communicative rather than diagnostic purpose. For example, if a student asked, "So you first got the headache 3 weeks ago?", some raters incorrectly scored this as an Investigative utterance (gathering new medical information) rather than an Interactional utterance (confirming understanding). These problems were solved with a redesign to produce JFHTA v9. The new v9 Part A (Interactional) consolidated v8's Part A2 and Part B, so all patient-centered and communicative element scoring appeared in a single section in v9, thus reducing the cognitive load for raters. Moving Part A Interactional to the top of the scoresheet made it more prominent, further reinforcing the improved usability by design. Additionally, in v9, the new Part B, Investigative, made it easier for raters to focus their attention here when a student asked a new question, intending to gather new medical information for diagnostic purposes. This fundamental redesign marked a major breakthrough, leading eventually to an Excellent ICC in v9.6.

The transition to a landscape orientation in later versions was an ergonomic necessity for digital assessment. In the Pilot phase, the portrait-oriented rubric (v8.3) placed "Undesirable Behaviors" (Part E) at the very bottom of the page. On a standard laptop screen, this required raters to scroll vertically to penalize negative behaviors. Consequently, while a rater was scrolling down to mark an inappropriate silence, they often missed a new utterance that occurred on screen. By aligning the digital scoresheet with the horizontal aspect ratio of a standard display, version 9.6 allowed "Undesirable Behaviors" (now moved up to Part C) to sit directly below the main scoring grids.

This layout change allowed raters to view the utterance tally boxes and the negative behavior checklist simultaneously. The feedback loop became immediate. Raters could mark negative behaviors the moment they occurred without navigating away from the primary scoring grid. Additionally, the integration of automated calculation via spreadsheet software eliminated arithmetic errors, ensuring that the variance in scores reflected true differences in student performance rather than calculation mistakes.

4.2 Theoretical foundations: quantifying patient-centeredness

JFHTA was designed to solve a specific assessment problem regarding the target cohort. The students are novice Japanese medical students with limited clinical knowledge and developing English proficiency. Traditional assessment tools often conflate linguistic accuracy with clinical competence. Our goal was to measure their ability to engage in a patient-centered interaction in a second language, with less weight placed on grammatical precision and fluency. To achieve this, the tool draws on the theoretical framework of Interactional Competence[8] and the methodological precedent set by the Roter Interaction Analysis System (RIAS).[7]

This approach aligns with the research of Cathy Heaven and colleagues who argued that patient-centeredness can be reliably measured if it is broken down into observable behaviors.[13] Following Heaven's precedent of quantifying "patient-centered" behaviors for assessment purposes, JFHTA moves away from subjective global ratings (e.g., "Was the student empathetic?") and towards a frequency-based model (e.g., "How many times did the student use an utterance categorized as empathy?"). By adopting a tally system similar to RIAS, JFHTA allows non-expert raters to objectively count the presence of patient-centered behaviors rather than subjectively judging their quality.

4.3 Structuring the encounter: the Interactional domain first approach

The structural reorganization of JFHTA v9.6 serves as a deliberate pedagogical signal. In version 8.3, "Medical Talk" (Part A1) was presented first, signaling to the student that diagnostic interrogation was the priority. In version 9.6, the Interactional section (Part A) is visually promoted to the top of the sheet. This forces a prioritization of communication and rapport building. Students in this context often gravitate towards reciting memorized lists of clinical questions without first establishing a human connection. By requiring the rater to assess communication and rapport first, the tool mirrors the logical flow of a real consultation: trust must be established before diagnostic interrogation begins. This redesign was also to foster the educational value of JFHTA as an assessment for learning, rather than an assessment of learning.

A specific challenge identified in the early phases was the high variance in how students opened the interview. Some students offered a single sentence before launching into medical questions, while others engaged in prolonged social chatting or struggled and asked for extensive clarification (spelling of names etc.) that consumed valuable time. In earlier JFHTA versions, raters struggled to score these variables consistently. Version 9.6 addressed this by introducing a dedicated "Greetings Section" within Part A.

This mechanism standardized the scoring of the beginning of the interview by offering a capped score of 7 points for specific obligatory behaviors: greetings, self-introduction, and patient identity confirmation (the Greetings Section). This creates a ceiling effect specifically for this Greeting Section, preventing artificial score inflation; students are rewarded for establishing safety and rapport, but they cannot inadvertently inflate their scores by chatting excessively before eliciting the chief concern. Importantly, Total Score variance was preserved (Range: 49–63), indicating that this local ceiling effect did not compress the overall discrimination capability of the tool. This reinforces the professional standard of efficient time management in clinical encounters.

4.4 Solving the non-clinician rater problem

One of the challenges in EMP assessment is the validity of scoring when the rater is not a physician. In Phase 1, the diverse rater panel struggled to agree on the quality of medical questions. For example, Rater 2 initially attempted to judge questions based on their perceived clinical diagnostic utility, while Rater 3 initially was distracted by grammatical inaccuracies. This discrepancy caused scoring errors which resulted in the low reliability scores seen in the Pilot phase.

JFHTA v9.6 addressed this by rigidly operationalizing the tool's original RIAS-inspired design. The revised criteria descriptors for Part B (Investigative) explicitly blocked the tendency for raters to erroneously drift into clinical judgment. Strictly defining and demanding 'evidence of listening' as part of the Investigative scoring criteria removed ambiguity and forced raters to align with the intended methodology. This effectively relieved the anxiety non-expert raters felt regarding the need to judge diagnostic value. By employing clear and strict definitions, v9.6 clarified that the rater's role is simply to count and categorize Investigative utterances based on the presence or absence of evidence of listening. This aligns with the concept of Interactional Competence, specifically the principle of contingency, which asserts that a competent speaker must actively attend to and build upon the contributions of their interlocutor.

This was operationalized through a two-tier scoring system. A standard medical question receives a baseline score of 1 point. However, to earn bonus points (3 points), the student must demonstrate concrete evidence of understanding by incorporating specific information provided by the patient into their follow-up question. For example, asking "When did it start?" is a generic question. Asking "When did the throbbing headache start?" demonstrates that the student actually understood the description provided by the patient. The final version of JFHTA explicitly disqualifies vague terms (e.g., "it," "that," "symptom") from receiving these bonus points, forcing a strict standard of evidence. This evidence-based follow-up question criterion allows non-clinician raters, including peers and language instructors, to reliably assess the "Medical" content of the history taking without requiring clinical expertise.

4.5 Feasibility and validity

The successful implementation of JFHTA confirms that it meets the practical demands of a large-scale medical curriculum. A key objective was to create a tool that was robust enough for summative assessment yet efficient enough for high-volume use. In many Japanese medical schools, a small number of instructors must evaluate over 100 students in a limited timeframe. A tool that requires complex medical decision-making or post-examination analysis is not feasible in this environment.

The robust ICC value of 0.92 in the final phase shows promise for summative use. It demonstrates that the specific mechanical changes successfully differentiated performance levels from grade S to D- without requiring a panel of expert judges. These results support the use of JFHTA as a programmatic assessment tool where data is collected routinely to track student growth over time. Furthermore, because the scoring criteria are based on observable behaviors rather than subjective clinical judgment, the tool is sufficiently transparent to be used for peer-assessment activities.

4.6 Limitations

This study has important limitations. First, the sample size of videos (n = 10) and raters (k = 3) was small, resulting in wide confidence intervals as detailed in Table 1. While the iterative process yielded high-point estimates of reliability (ICC = 0.92), the wide intervals indicate low statistical precision; therefore, these results should be interpreted as a pilot validation of the development process rather than a definitive verification.

Second, Rater 1 was the developer of the tool and the course instructor, meaning they were unblinded to the students' identities. This introduces a risk of bias where the improvement in reliability may be attributed to Rater 1 simply training Raters 2 and 3 during the "Check" phases. It is possible that this process acted as calibration training, creating a 'shared mental model' among this specific group of raters. However, it is notable that reliability did not improve linearly simply due to increased rater familiarity; Phase 2 (v9.3) and Phase 3 (v9.5) showed persistent disagreement in the Interactional Scores despite the raters having already collaborated in Phase 1. The jump to "Excellent" reliability in the Interactional Score occurred in Phase 4, resulting from the refined wording, the reorganization of the behavioral frequency counts and overall design improvements. The JFHTA scoresheet/rubric design, rather than mere rater calibration, played a decisive role.

Third, the study was conducted at a single private medical university with a specific curriculum. The generalizability of JFHTA to other EMP programs remains to be tested. Finally, the tool was designed specifically for novice learners of English interacting with simulated patients. Its sensitivity may differ when applied to different cohorts of students.

4.7 Conclusion

Improving JFHTA's reliability was not only a product of rater training. It resulted from collaborative improvements to instrument design . While earlier JFHTA versions were developed largely in isolation, the improved consistency of v9.6 was mainly the result of the PDCA process. The diverse team of experts identified and resolved sources of ambiguity over multiple cycles, which transformed an unreliable tool into a more robust assessment method with potential utility beyond the host institution's curriculum. The innovations in JFHTA offer a model for other institutions seeking to assess clinical communication skills in a second language. By reducing the cognitive load on raters and defining a clear method for counting and categorizing utterances, we can begin to enable non-clinical educators to assess future doctors with greater precision and confidence.

Acknowledgments

We acknowledge the use of the host institution's licensed instance of Gemini Pro for assistance with linguistic refinement and data formatting during manuscript preparation.

References

1. Hoshina Y, Yada K, Maki H, Yoshino T, Takaiso H, and Akaike M. 2022. Medical English education in Japan: Developing a curriculum to motivate students by providing visualization opportunities using near-peer teaching. *J Med Invest* **69**(3-4): 332–334.
2. Medical Institutions Increasingly Handling Foreign Patients; Areas Responding to Increase in Overseas Visitors, Residents - The Japan News [Internet]. [cited Jan. 7, 2026]. Available from: https://japannews.yomiuri.co.jp/society/general-news/20250520-255499/
3. Expanding Medical Tourism to Japan for the Mutual Benefit of Japan and the World ｜国際経済交流財団 (JEF: Japan Economic Foundation) [Internet]. [cited Jan. 7, 2026]. Available from: https://www.jef.or.jp/cmb/expanding_medical_tourism_to_japan_for_the_mutual_benefit_of_japan_and_the_world.html
4. Fujikawa H, Tamune H, Nishizaki Y, Shikino K, Shimizu T, Yamamoto Y, et al. 2025. Undergraduate general medicine education in Japan: A nationwide cross-sectional survey of medical trainees' perspectives. *J Gen Fam Med* **26**(2): 148–156.
5. Asada Y. 2024. Learning Objectives from the Syllabus and Model Core Curriculum for Medical Education in Japan: A Case Report of Mapping. IIAI Lett Institutional Res [Internet]. Sept. 15, 2024 [cited Jan. 8, 2026]. Available from: https://iaiai.org/letters/index.php/lir/article/view/295
6. Jego EH. 2022. Assessing oral communication skills of EFL novice medical students: interactional competence and the Roter Interaction Analysis System [PhD thesis]. University of Birmingham [cited Jan. 8, 2026]. Available from: https://etheses.bham.ac.uk/id/eprint/12615/
7. Roter D and Larson S. 2002. The Roter interaction analysis system (RIAS): Utility and flexibility for analysis of medical interactions. *Patient Educ Couns* **46**(4): 243–251.
8. Young RF. 2011. *Interactional Competence in Language Learning, Teaching, and Testing*. In: Hinkel E, editor. Handbook of Research in Second Language Teaching and Learning. Vol. 2. Routledge: p. 426–443.
9. Nara N. 2025. Impact of accreditation on medical education quality improvement in 82 medical schools in Japan: A descriptive study. *J Educ Eval Health Prof* **22:** 22.
10. Shrout PE, Fleiss JL. 1979. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull* **86**(2): 420–428.
11. McGraw KO and Wong SP. 1996. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* **1**(1): 30–46.
12. Koo TK and Li MY. 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* **15**(2): 155–163.
13. Heaven C, Maguire P, and Green C. 2003. A patient-centred approach to defining and assessing interviewing competency. *Epidemiol Psichiatr Soc* **12**(2): 86–91.

Conference proceedings

Health screening tourism: Toward improving hospitality for foreign tourists in Osaka

Tsuyoshi Sakoda and Koshiro Ando
ISEIKAI International General Hospital
Sophia Complete Medical Checkup Center

Abstract

Osaka is experiencing growth in health screening tourism, attracting more foreign visitors. However, many medical institutions operate mainly in Japanese, posing challenges due to language and cultural barriers. Common issues include difficulty understanding medical terms, limited staff language skills, and differing cultural attitudes toward care.

To improve service quality, several strategies are proposed: using visual aids to assist understanding, employing digital translation tools for basic communication, offering cross-cultural training to staff, simplifying medical terminology, and proactively providing English explanations. In one case, a Chinese tourist better understood results when explained in English by a doctor than through a Chinese interpreter, suggesting that some affluent Chinese tourists may prefer English communication. On the other hand, a Chinese couple, accompanied by a friend with limited Japanese skills, struggled during reception and checkup explanations, requiring translation AI that was time-consuming and inefficient.

Health screening tourism encompasses more than medical treatment—it includes the overall travel and service experience. Ensuring smooth communication, high-quality care, and comfortable accommodations is essential to encourage repeat visits and positive referrals. By implementing these strategies, Osaka's healthcare providers can enhance service reliability and contribute to the sustainable growth of medical tourism.

J Med Eng Educ (2026) 25 (1): 30-35

Keywords health screening tourism, Osaka, hospitality, language barriers, cultural differences, medical English

1. Introduction

Osaka, a vibrant metropolitan hub in Japan, has experienced a remarkable surge in international tourism in recent years. This growth extends beyond leisure and business travel to encompass health and medical tourism, with a notable increase in foreign visitors seeking comprehensive health screenings and advanced medical treatments. The appeal of Japan's advanced medical technologies, high standards of care, and renowned hospitality has positioned Osaka as an attractive destination for individuals seeking preventive medicine and diagnostic services. However, this burgeoning sector presents unique challenges, particularly concerning the provision of medical services to a diverse international clientele whose primary language is often not Japanese. Language and cultural disparities can significantly impede effective communication between medical professionals and foreign patients, potentially affecting the quality of care, patient satisfaction, and overall hospitality experience.[1]

The core problem addressed in this paper is the need to enhance hospitality for foreign tourists engaging in health screening tourism in Osaka, specifically within medical institutions that predominantly operate in Japanese and may have limited multilingual support. The objective of this research is to identify the key challenges posed by language and cultural barriers in this context and to propose effective strategies for medical institutions to improve their reception systems and service delivery, ultimately fostering a more reliable, comfortable, and satisfying experience for international health screening tourists. By doing so, Osaka can solidify its position as a leading destination for medical tourism and ensure the sustainable growth of this vital sector.

Corresponding author:
Tsuyoshi Sakoda
M.D. Ph.D., ISEIKAI International General Hospital, 4-14 Minamiogimachi, Kita-ku, Osaka-shi, Osaka, 530-0052 Japan
TEL: 0570-099166
E-mail: sakoda@msd.biglobe.ne.jp

Presented at: Japan Society for Medical English Education 2025 Annual Conference, Osaka, Japan

Table 1. Number of international visitors to Osaka (in millions)

Year	Visitors (in millions)
2019	31.8
2023	25
2025 (Goal)	40
2030 (Goal)	60

2. Rapid growth of inbound tourism in Osaka

Osaka has witnessed a remarkable increase in inbound tourism over the past decade. In 2023, the city welcomed over 25 million international visitors, closely approaching its pre-pandemic peak of 31.8 million in 2019. This impressive recovery and continued growth underscore Osaka's appeal as a global destination. The city has set ambitious targets for future growth, aiming to attract 40 million international visitors at the end of 2025 and an even more ambitious 60 million by 2030. These targets are significantly driven by major upcoming events and developments, notably Expo 2025 and the integrated resort (IR) development, both of which are expected to further boost visitor numbers and stimulate the local economy.[2-5]

Table 1 shows the historical and projected growth of international visitors to Osaka. This sustained growth in inbound tourism provides a fertile ground for the expansion of specialized tourism sectors, including health and medical tourism. The increasing influx of international visitors, coupled with their rising spending capacity, creates a significant market for high-quality health screening and medical services.

3. Osaka's medical tourism market

Within Osaka's rapidly expanding tourism landscape, there is a growing and discernible demand for health and medical tourism. This demand is particularly pronounced for advanced health screenings and specialized medical treatments. Major hospitals in Osaka are already experiencing this trend firsthand, with many receiving hundreds of international patients annually. These patients seek a range of services, including comprehensive health checkups, sophisticated cancer screenings, and cutting-edge advanced therapies. The reputation of Japanese medicine for precision, technological advancement, and patient-centered care makes it a highly attractive option for medical tourists globally.

Recognizing the growing importance of this sector, the Osaka Convention & Tourism Bureau, in collaboration with JTB, took a significant step in 2023 by launching the "Osaka International Medical Contribution Platform". This initiative aims to strengthen the support infrastructure for international medical tourists, streamlining their access to medical services and ensuring a more coordinated and seamless experience.[6]

An analysis of the country breakdown of health screening tourists reveals a dominant trend: over 90% of these visitors originate from China.[1] This highlights the significant market potential within China for medical tourism in Osaka. While China remains the primary source, this diversification, though nascent, indicates a broader appeal of Osaka's medical services across the Asian region. Understanding these demographic trends is crucial for tailoring services and marketing efforts to meet the specific needs and preferences of different patient groups.

4. Osaka's healthcare reception system

In response to the increasing influx of foreign patients, major hospitals in Osaka are observing a steady rise in the number of international patients seeking medical services. To effectively manage this growing demand and enhance the overall experience for medical tourists, the newly established "Osaka International Medical Contribution Platform" plays a central role. This platform is designed to centralize various aspects of patient reception, offering comprehensive multilingual support and facilitating efficient care coordination.[6]

The enhanced support provided through this platform and by individual medical institutions includes a range of services crucial for international patients. These services go beyond mere medical treatment to encompass logistical and administrative support, such as making stay arrangements, facilitating payment processes, and even assisting with medical visa sponsorship. By providing such holistic support, Osaka aims to create a welcoming and convenient environment for medical tourists, addressing many of the non-medical concerns that can arise when seeking healthcare abroad. The centralization of these services through the new platform is a critical step towards creating a more streamlined and patient-friendly healthcare reception system for international visitors in Osaka.

5. The appeal of health screening tourism

Osaka's growing prominence in health screening tourism is underpinned by several compelling factors that attract international visitors. Foremost among these are the high reputation of Japan's preventive medicine and its advanced diagnostic capabilities. Japanese medical institutions are renowned for their expertise in sophisticated diagnostic procedures, including state-of-the-art endoscopy, advanced imaging techniques, and cutting-edge therapies such as immunotherapy and gene therapy. These advanced technologies and the precision with which they are applied offer a significant appeal to individuals seeking thorough and accurate health assessments.

Beyond the technological prowess, the overall patient

experience in Japanese healthcare settings is a major attraction. Foreign tourists are often drawn by the meticulously clean facilities, the attentive and meticulous care provided by medical staff, and the overarching embodiment of Japanese hospitality, known as "omotenashi". This focus on cleanliness, patient comfort, and a high level of service contributes significantly to a positive and reassuring medical experience. The combination of world-class medical technology and a culturally ingrained commitment to patient well-being makes Osaka a highly desirable destination for health screening tourism.

6. Challenges: Language and cultural barriers

Despite the growing appeal and advancements in Osaka's medical tourism sector, significant challenges persist, primarily centered around language and cultural barriers. These barriers can profoundly impact the quality of communication and the overall patient experience.

6.1 Language gaps

The most immediate and apparent challenge is the language gap. Foreign patients often struggle to understand complex medical explanations and terminology when presented in Japanese. This lack of understanding can lead to anxiety, confusion, and a diminished ability for patients to make informed decisions about their health. Even basic communication can be hindered, creating inefficiencies in the registration process, during consultations, and throughout the treatment journey.[7]

6.2 Cultural differences

Beyond language, cultural differences play a crucial role in shaping patient expectations and attitudes toward care. What might be considered standard medical practice or communication style in Japan may differ significantly from a patient's home country. This can lead to misunderstandings regarding appointment protocols, the level of detail in explanations, the role of family in decision-making, and even expressions of pain or discomfort. For instance, some cultures may prefer direct and blunt communication, while others may value indirectness and restraint, potentially leading to misinterpretations if cultural nuances are not understood.

6.3 Staffing and technological limitations

A contributing factor to these communication issues is the shortage of multilingual staff within medical institutions, especially with regard to medical English. Morita et. al showed that even if foreign residents have no trouble speaking Japanese in their daily lives, they do not always understand the language spoken in hospitals.[8] This is because they face many barriers other than language. While some larger hospitals may have dedicated international departments, many smaller clinics and even some departments within larger hospitals may not have sufficient staff proficient in languages other than Japanese. Furthermore, while digital translation tools and artificial intelligence (AI) translation have emerged as potential solutions, they come with their own limitations. AI translations can be time-consuming, sometimes inaccurate, and may fail to capture the nuances and subtleties of medical terminology or cultural context, leading to inefficient communication and potential misunderstandings.

6.4 Communication barriers in practice

Real-world scenarios vividly illustrate these communication challenges, as shown in **Table 2**:

- **English explanation:** When medical explanations were provided in English to tourists with strong English skills, the communication was well-received, leading to direct understanding and higher patient satisfaction. This highlights the effectiveness of direct communication when language proficiency exists.
- **Friend as interpreter:** In one instance, a Chinese couple accompanied by a friend who claimed to speak Japanese encountered significant difficulties at the reception desk due to the friend›s insufficient Japanese language skills. This scenario demonstrates the unreliability of informal interpreters, who may lack the medical vocabulary or the ability to accurately convey complex information.
- **AI translation:** For patients who understood neither English nor Japanese, the reliance on AI translation during results explanations proved to be time-consuming and labor-intensive. While helpful for basic phrases, the limitations of AI in conveying intricate medical details were evident, leading to potential misunderstandings and inefficiencies.

These practical examples underscore the critical need for robust and reliable communication strategies to bridge the linguistic and cultural divides in health screening tourism.

Table 2. Communication barriers in practice

Scenario	Problem	Outcome
English Explanation	Well received by tourists with English skills	Direct communication, higher satisfaction
Friend as Interpreter	Friend's Japanese is insufficient	Confusion at reception
AI Translation	Time-consuming, sometimes inaccurate	Inefficient, prone to misunderstanding

7. Solutions at medical institutions

Addressing the multifaceted challenges of language and cultural barriers requires a proactive and comprehensive approach from medical institutions in Osaka. Several key solutions can be implemented to significantly enhance the hospitality and quality of care for foreign health screening tourists.

7.1 Introduction of visual aids in medical explanations

One highly effective strategy is the increased introduction and systematic use of visual aids during medical explanations. Visuals, such as anatomical diagrams, simplified flowcharts of procedures, and images from diagnostic tests (e.g., MRI, endoscopy results), can transcend language barriers by providing clear, concrete representations of information. For example, using images from anatomical figures from textbooks to explain human anatomy or displaying results of scans can greatly aid understanding for non-Japanese speakers. This approach allows medical professionals to point to specific areas of concern, illustrate complex processes, and reinforce verbal explanations, even if those explanations are delivered through an interpreter or a translation tool.

7.2 Judicious use of digital translation tools

Digital translation tools can be helpful for basic communication and routine inquiries, particularly at reception desks or for simple instructions. However, their use should be approached with caution, especially for critical medical discussions. Staff should be trained on the limitations of these tools, understanding that they are supplementary aids rather than replacements for human linguistic expertise. For more nuanced and complex medical conversations, direct human interpretation or communication is highly preferable to minimize the risk of misinterpretation.

7.3 Centralized patient support through medical tourism platform

The "Osaka International Medical Contribution Platform" provides a crucial infrastructure for centralized patient support. Medical institutions should leverage this platform to its full potential for managing appointments, coordinating care across different departments, facilitating payment processes, and assisting with logistical aspects such as accommodation and transportation. A centralized system ensures consistency in service delivery and provides a single point of contact for international patients, reducing confusion and streamlining their medical journey.

7.4 Proactive offering of English language services

Given the proficiency in English of many affluent Chinese clients and their preference for direct communication, medical institutions should proactively offer and promote English-language services. This includes having English-speaking staff available, providing medical information brochures in English, and developing English versions of essential forms and consent documents. Emphasizing direct communication in English, especially for detailed medical explanations, can significantly enhance patient satisfaction and understanding, as demonstrated by the case of the Chinese neurosurgeon described below in "Case studies".

7.5 Cross-cultural training for staff

To address cultural differences affecting patient expectations and attitudes, medical staff should receive cross-cultural training. This training should aim to raise awareness of diverse cultural norms related to healthcare, communication styles, privacy, and decision-making processes. Understanding these differences can help staff anticipate patient needs, avoid misunderstandings, and provide culturally sensitive care that respects the patient's background.

7.6 Simplification of medical terminology

Regardless of the language of communication, medical institutions should strive to simplify medical terminology when explaining conditions, diagnoses, and treatment plans to patients. Using plain language, avoiding jargon, and employing analogies can make complex medical information more accessible. This approach benefits all patients, but it is particularly critical for those who are navigating a healthcare system in a foreign language.

By systematically implementing these solutions, medical institutions in Osaka can create a more welcoming, efficient, and reliable environment for international health screening tourists, thereby enhancing their overall experience and fostering positive word-of-mouth referrals.

8. Case studies

Practical scenarios underscore the varying effectiveness of different communication approaches and highlight the critical need for robust language support in health screening tourism.

8.1 English communication with wealthy Chinese clients

A notable observation in Osaka's medical tourism sector is that affluent Chinese clients frequently possess strong English language skills and often express a preference for direct explanations in English from medical professionals. This preference for direct doctor-patient communication in English significantly reduces the potential for misunderstandings that can arise when interpreters are used. In one specific instance, a thorough discussion between a Chinese client and our staff was conducted entirely in

English, leading to remarkably high client satisfaction. This experience served as a compelling demonstration of the necessity for medical staff to acquire and utilize English language skills when attending to international clients. The ability to engage in direct, in-depth conversations fosters trust and clarity, which are paramount in healthcare settings.

8.2 Medical English with a Chinese neurosurgeon

A particularly insightful case involved a results explanation session with a client who was subsequently revealed to be a neurosurgeon, and who presented with a cerebral aneurysm. Doctors in both China and the US had previously recommended surgery for his condition. However, the Japanese doctor suggested conservative medical therapy, which the client ultimately chose.

During the explanation of his health check-up results in our hospital, MRI findings indicated that surgery could still be considered, so I recommended another neurological consultation to discuss future treatment options. Despite the in-depth discussion about the aneurysm, the client's final decision was to continue with medication. Crucially, this complex medical discussion was conducted entirely in English. This experience profoundly emphasized that many wealthy Chinese clients, particularly those with a medical background like this neurosurgeon, require not only general English proficiency[9] but also specialized medical English skills. This case strongly reinforced the importance of continuous medical English education and training for hospital staff to effectively serve the needs of a sophisticated international clientele.

These case studies collectively demonstrate that while the demand for medical tourism is growing, the quality of the patient experience is inextricably linked to the effectiveness of communication. Direct English communication is often preferred by and more effective for certain demographics, while a lack of professional language support can lead to significant inefficiencies and potential misunderstandings, reinforcing the need for targeted linguistic and cultural competencies within medical institutions.

9. Discussion

The rapid growth of inbound tourism in Osaka presents both immense opportunities and significant challenges for its healthcare sector. The increasing demand for health screening tourism underscores Osaka's potential as a global medical hub, leveraging its reputation for advanced diagnostics and preventive medicine, coupled with the distinctive appeal of Japanese hospitality. However, realizing this potential for sustainable growth hinges on effectively overcoming existing language and cultural barriers that currently impede optimal patient care and satisfaction.

The case studies presented vividly illustrate the complexities of cross-cultural medical communication. The preference of affluent Chinese clients for direct English explanations from doctors, as opposed to interpreter-mediated communication, highlights a crucial demographic need. This is not merely about general English proficiency but extends to specialized medical English, as demonstrated by the interaction with the Chinese neurosurgeon. For medical institutions, this implies a strategic imperative: investing in the medical English proficiency of their staff is no longer a luxury but a fundamental requirement for delivering high-quality care to this influential segment of medical tourists. The direct communication fostered through shared linguistic competence reduces the risks of misunderstanding and enhances patient trust and satisfaction.

Conversely, the challenges encountered when relying on informal interpreters or basic AI translation tools underscore their limitations for nuanced medical discussions. While these tools may serve for rudimentary exchanges, the intricacies of medical terminology and the importance of accurate information conveyance necessitate more sophisticated solutions. This reinforces the argument for professional interpretation services or, ideally, direct communication from medically trained, multilingual staff. The time-consuming and often inaccurate nature of AI translation for complex scenarios can lead to inefficiencies and potentially compromise patient safety and understanding.

The proposed solutions, including the introduction of visual aids, the judicious use of digital tools, centralized patient support, proactive English services, cross-cultural training, and simplification of medical terminology, offer a multi-pronged approach to enhance the patient experience. Visual aids, in particular, provide a universal language that can bridge gaps where verbal communication falters. The "Osaka International Medical Contribution Platform" serves as a vital piece of infrastructure for coordinating these efforts, ensuring a seamless journey for medical tourists from pre-arrival to post-treatment.

Ultimately, improving hospitality in health screening tourism extends beyond just clinical excellence. It encompasses the entire patient journey, from initial inquiry to follow-up care. Cultural sensitivity, efficient administrative processes, and robust language support are all integral components of a truly hospitable environment. By addressing these aspects comprehensively, Osaka's medical institutions can not only meet the immediate needs of their international patients but also build a reputation that encourages repeat visits and positive word-of-mouth referrals, essential for long-term sustainable growth.

10. Conclusion

The rapid growth of international tourism in Osaka, coupled with ambitious targets for future visitor numbers

driven by Expo 2025 and IR development, presents a significant opportunity for the expansion of health screening tourism. Osaka's appeal is rooted in its high reputation for preventive medicine, advanced diagnostics, clean facilities, attentive care, and quintessential Japanese hospitality. However, realizing the full potential of this sector necessitates a concerted effort to overcome persistent language and cultural barriers that currently challenge the provision of quality medical services to foreign tourists.

The findings emphasize that effective communication is paramount for patient satisfaction and successful medical outcomes. While the preference for direct English communication among affluent Chinese clients highlights the value of medical English proficiency among staff, the limitations of informal interpreters and current AI translation tools for complex medical discussions underscore the need for more reliable and professional language support.

To enhance hospitality and ensure sustainable growth in health screening tourism, medical institutions in Osaka must proactively implement comprehensive strategies. These include the widespread use of visual aids, the cautious and informed application of digital translation tools, leveraging the centralized support offered by the "Osaka International Medical Contribution Platform," and a strong commitment to providing English-language services. Furthermore, ongoing cross-cultural training for staff and the simplification of medical terminology are essential to address the nuanced aspects of patient expectations and understanding.

By addressing these challenges comprehensively, Osaka can significantly improve the reliability and quality of its health screening services for international tourists. This, in turn, will not only contribute to the sustainable growth of medical tourism but also reinforce Osaka's standing as a welcoming and world-class destination for global health and wellness. Strengthened collaboration among hospitals, tourism bodies, and government sectors is crucial to constructing a safe and reliable system that can meet the evolving demands of this vital sector.

11. Prospects for sustainable growth

The future of health screening tourism in Osaka holds significant promise, contingent upon strategic and collaborative efforts. For sustainable growth, it is imperative to establish and strengthen collaboration among hospitals, the broader tourism industry, and government sectors. This multi-stakeholder cooperation is essential for constructing a safe, reliable, and integrated system that can seamlessly cater to the needs of international medical tourists. Such collaboration could involve joint marketing initiatives, standardized service protocols, and shared resources for language and cultural support.

A key driver for repeat visits and positive word-of-mouth referrals, which are vital for long-term success, will be the continuous enhancement of language and cultural support within medical institutions. This goes beyond merely providing translation services to fostering a deeply understanding and respectful environment for diverse international patients. Ongoing training, cultural competency programs, and the recruitment of multilingual staff will be crucial investments in this regard.

At present, major events such as Expo 2025 are underway and the planned Integrated Resort (IR) development are poised to significantly boost Osaka's medical tourism sector. These developments will not only increase the overall influx of international visitors but also elevate Osaka's global profile, attracting more individuals seeking high-quality health and medical services. By proactively preparing and refining its healthcare reception systems and hospitality standards, Osaka can fully capitalize on these opportunities, solidifying its position as a premier destination for health screening tourism globally.

References

1. nippon.com. 2023. December 15. Popular health checkups and advanced medical care: The evolution of Japan's 'medical tourism' as conveyed by pioneer. JTB. In Japanese.
2. Sumitomo Mitsui Trust Asset Management Co., Ltd. 2020. July 6. Recovery in Japanese Tourism.
3. Tochibayashi N. 2025. How Japan is redesigning tourism to benefit local communities. *World Economic Forum* Nov. 18.
4. y-NEX. 2024 December 13. Can Japan become a powerhouse in tourism? December 13. 2024.
5. Allen J. 2025. January 10 Report: Japan Will See Record 40 Million Tourists in 2025. UNSEEN JAPAN.
6. JTB. 2023, March 20. Establishment of "Osaka International Medical Contribution Platform" with Osaka Convention & Tourism Bureau by "Osaka International Medical Contribution Promotion Committee." In Japanese.
7. Irimajiri S. 2025. Increase in the Number of Foreign Visitors to Japan and the Importance of English Communication Skills. *J Med Eng Educ* **24**(2): 22-23.
8. Morita N and Yoshitomi S. 2020. Expectations and proposals from the healthcare interpreter's perspective. *Medical Education* **51**(6): 643-649.
9. Willey I, Tanimoto K, McCrohan G, and Nishiya K. 2020. An English needs analysis of medical doctors in Western Japan. *JALT Journal* **42**(2): 143-169.