Dovepress

Open Access Full Text Article

ORIGINAL RESEARCH

Validation of the General Medicine in-Training Examination Using the Professional and Linguistic Assessments Board Examination Among Postgraduate Residents in Japan

Kazuya Nagasaki[1]
Yuji Nishizaki[2,3]
Masanori Nojima[4]
Taro Shimizu[5]
Ryota Konishi [6]
Tomoya Okubo[7]
Yu Yamamoto[8]
Ryo Morishima[9]
Hiroyuki Kobayashi[1]
Yasuharu Tokuda [10]

[1]Department of Internal Medicine, Mito Kyodo General Hospital, University of Tsukuba, Ibaraki, Japan; [2]Medical Technology Innovation Center, Juntendo University, Tokyo, Japan; [3]Division of Medical Education, Juntendo University School of Medicine, Tokyo, Japan; [4]Center for Translational Research, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan; [5]Department of Diagnostic and Generalist Medicine, Dokkyo Medical University Hospital, Tochigi, Japan; [6]Education Adviser Japan Organization of Occupational Health and Safety, Kanagawa, Japan; [7]Research Division, The National Center for University Entrance Examinations, Tokyo, Japan; [8]Division of General Medicine, Center for Community Medicine, Jichi Medical University School of Medicine, Tochigi, Japan; [9]Department of Neurology, Tokyo Metropolitan Neurological Hospital, Tokyo, Japan; [10]Muribushi Okinawa for Teaching Hospitals, Okinawa, Japan

Correspondence: Yuji Nishizaki
Medical Technology Innovation Center, Juntendo University, 2-1-1 Hongo Bunkyo-ku, Tokyo, 113-8421, Japan
Tel +81-3-3813-3111
Fax +81-3-5689-0627
Email ynishiza@juntendo.ac.jp

Purpose: In Japan, the General Medicine In-training Examination (GM-ITE) was developed by a non-profit organization in 2012. The GM-ITE aimed to assess the general clinical knowledge among residents and to improve the training programs; however, it has not been sufficiently validated and is not used for high-stake decision-making. This study examined the association between GM-ITE and another test measure, the Professional and Linguistic Assessments Board (PLAB) 1 examination.

Methods: Ninety-seven residents who completed the GM-ITE in fiscal year 2019 were recruited and took the PLAB 1 examination in Japanese. The association between two tests was assessed using the Pearson product-moment statistics. The discrimination indexes were also assessed for each question.

Results: A total of 91 residents at 17 teaching hospitals were finally included in the analysis, of whom 69 (75.8%) were women and 59 (64.8%) were postgraduate second year residents. All the participants were affiliated with community hospitals. Positive correlations were demonstrated between the GM-ITE and the PLAB scores ($r = 0.58$, $p < 0.001$). The correlations between the PLAB score and the scores in GM-ITE categories were as follows: symptomatology/clinical reasoning ($r = 0.54$, $p < 0.001$), physical examination/procedure ($r = 0.38$, $p < 0.001$), medical interview/professionalism ($r = 0.25$, $p < 0.001$), and disease knowledge ($r = 0.36$, $p < 0.001$). The mean discrimination index of each question of the GM-ITE (mean ± SD; 0.23 ± 0.15) was higher than that of the PLAB (0.16 ± 0.16; $p = 0.004$).

Conclusion: This study demonstrates incremental validity evidence of the GM-ITE to assess the clinical knowledge acquisition. The results indicate that GM-ITE can be widely used to improve resident education in Japan.

Keywords: in-training examination, validity, extrapolation, General Medicine In-Training Examination, Professional and Linguistic Assessments Board, medical knowledge, postgraduate medical education

Introduction

In Japan, medical students enter a 2-year postgraduate residency program after completing their medical school education.[1,2] They are referred to as "postgraduate residents" or simply "residents." This training period aims to help residents acquire basic clinical knowledge and general medical skills. Residents are required to rotate through seven specialties (internal medicine, emergency medicine, community medicine, surgery, anesthesiology, pediatrics, psychiatry, and obstetrics and gynecology) during this

2-year period and receive supervised training. After that, most residents enter a specialty-based residency training. The General Medicine In-training Examination (GM-ITE) is an in-training examination of clinical knowledge during the initial two-year postgraduate residency program. The GM-ITE was developed by the Japan Organization of Advancing Medical Education (JAMEP; a non-profit organization) in 2011 by using a methodology similar to that of the US Internal Medicine Residency Examination (IM-ITE).[3] The purpose of GM-ITE is to facilitate the improvement of training programs by providing residents and program directors with an objective, reliable assessment of clinical knowledge.[4] The examination contains 60 multiple-choice questions (MCQs) covering a wide range of clinical knowledge, from physical examinations and procedures to diagnosis, treatment, and psychosocial care. The GM-ITE is not used as a pass/fail for training advancement, without a cutoff for the exam. On the other hand, the Ministry of Health, Labor and Welfare (MHLW) is considering making this exam mandatory for all residents and use it to improve clinical training. The examination is voluntary for each training hospital, and approximately one-third of residents take the examination each year (6133 residents in fiscal year 2018, 5593 residents in fiscal year 2017, and 4568 residents in fiscal year 2016).[5]

The validation of the GM-ITE was not well assessed in the literature. Validation is the process of collecting and interpreting evidence to demonstrate the accuracy of inference drawn from the test performance. Kane demonstrates a framework to collect evidence of inference with four components: scoring, generalization, extrapolation, and interpretation (or decision).[6] The JAMEP organizes a question development committee of experienced physicians in various fields. The quality of the question was preserved by an independent peer-review committee and experts in examination analysis, which provide sufficient evidence to scoring component. The GM-ITE is based on the content of the objectives of the clinical training presented by the MHLW.[7] The MHLW also requires residents to master professionalism; physical examination and clinical procedure; and the diagnosis and treatment of common diseases. As the examination covers four categories, including clinical reasoning, physical examination/procedure, communication/professionalism, and disease knowledge, the GM-ITE has evidence of generalization component; however, the examination contains a relatively small number of questions, suggesting low generalization evidence. To strengthen the evidence on extrapolation, the GM-ITE uses the clinical vignette

format in most questions. The exam also includes video and audio questions to approximate real-world situations. The GM-ITE is not currently used for any decision-making about residents or residency programs and the evidence for the interpretation component cannot be evaluated. The JAMEP has been working to increase its validity and to promote its use in higher-stake decision-making in postgraduate clinical training.

The validity of tests with multiple-choice questions often has problems with the evidence during extrapolation.[8] A method that examines extrapolability should investigate the relationship between actual clinical performance and test performance; however, general clinical skills in actual clinical practice cannot be measured. Therefore, we decided to examine the relationship between the GM-ITE and other tests with similar purposes of acquiring knowledge of general medical skills. The comparator examination in this study was selected on the basis of its similarity to the GM-ITE and its availability. During the selection process for the exam, we also emphasized the fact that it includes a wide range of questions and that some examinees came from other countries. Finally, we decided to use the Professional and Linguistic Assessments Board (PLAB).[9] This examination is designed for the purpose of proving that International Medical Graduates have the necessary skills and knowledge to practice medicine in the United Kingdom. It is designed to assess the depth of medical knowledge and levels of medical and communication skills. The PLAB examination has two parts: part 1 is composed of MCQs, and part 2 is an objective structured clinical examination. As the GM-ITE only contains MCQs, we use the PLAB 1 examination for this study.

The evaluation of the residency program by in-training examination is important for improving the educational environment of postgraduate medical residents.[10] Further validation is needed to utilize the results of the GM-ITE to improve the residency program in Japan. The aim of this study was to assess the association between the GM-ITE using the PLAB 1 examination.

Methods
Participants, Data Collection, and Measurement

We conducted a multicenter cross-sectional validation study of PGY-1 and PGY-2 residents in Japan. We recruited 97 participants of 17 teaching hospitals from among the examinees of the GM-ITE 2019, which was administered in January to February 2020. After obtaining

their consents, we administered the PLAB 1 examination to the study participants. We collected the scores in both examinations and the characteristic of the participants. We excluded from the analysis the participants who did not complete the PLAB 1 examination. All the participants read and signed the informed consent document before the survey. The ethics review board of Juntendo University School of Medicine approved the study.

Participants

Both the GM-ITE and the PLAB 1 examinations consist of MCQs. The GM-ITE contains 60 MCQs; however, one question (question 34) was excluded from the analysis because the question was considered inappropriate in the post-examination evaluation process.

With the permission of the General Medicine Council (GMC) in the UK, only the multiple-choice written exam questions (PLAB part 1) were used. First, the translatio-nand validation processes involved translating the 200 questions from English to Japanese, and back-translating from 150 Japanese to English.[11] During this process, 20 questions were excluded because they were irrelevant to the Japanese medical setting. Then, three Japanese physicians (two internists and one emergency physician) inspected the content of each question and adjusted the items to reflect the situation in the Japanese medical field. Finally, the Japanese version contains 180 MCQs. All residents took the PLAB 1 examination in Japanese.

Statistical Analyses

The association between the GM-ITE and PLAB 1 examination scores was assessed using the Pearson product-moment statistics (Pearson correlation coefficient = r), where r ≥ 0.40 was considered satisfactory. Correlation coefficients were assessed between the total and subcategory scores of the GM-ITE and PLAB 1 examination. Correlations were considered significant at the 0.01 level (two-tailed). The discrimination indexes of the questions in the GM-ITE and PLAB 1 examination were calculated using the correctness for the top 25% and bottom 25% of the rankings.[12,13] A discrimination index of ≥0.20 indicates that the question has a high discriminatory power, and a discrimination index of >0.40 indicates that the question is a very good measure of the subject's qualifications, assuming that variability exists in the group. The mean discrimination indexes of both examinations were assessed using a t test. We set $p < 0.05$ as statistically

significant. All analyses were conducted using SPSS 25 (IBM, Armonk, NY, USA).

Results

Ninety-seven residents completed the GM-ITE and consented to participate in the study. Excluding six residents who did not complete the PLAB 1 examination, 91 residents at 17 teaching hospitals were finally included in the analysis. Among the study participants, 69 (75.8%) were women and 59 (64.8%) were PGY-2 residents. One resident had previously taken the PLAB examination. All the participants were affiliated with community hospitals. The mean (±SD) GM-ITE score was 30.2 ± 5.5, and the mean PLAB 1 examination score was 119.5 ± 11.7 (Table 1).

The scatterplot of the GM-ITE and PLAB 1 examination scores are shown in Figure 1. Positive correlations were demonstrated between the total scores of the GM-ITE and PLAB 1 examination ($r = 0.58$, $p < 0.001$). The correlations between the scores in the PLAB 1 examination and GM-ITE categories were as follows: symptomatology/clinical reasoning ($r = 0.54$, $p < 0.001$), physical examination/procedure ($r = 0.38$, $p < 0.001$), medical interview/professionalism ($r = 0.25$, $p < 0.001$), and disease knowledge ($r = 0.36$, $p < 0.001$; Table 2).

The discrimination indexes of the GM-ITE and the PLAB 1 examination scores are demonstrated in Figure 2. The mean discrimination indexes of the questions in the GM-ITE (mean ± SD; 0.23 ± 0.15) was higher than those of the questions in the PLAB (0.16 ± 0.16; $p = 0.004$; Appendix 1).

Discussion

We found a strong positive correlation between the GM-ITE and PLAB 1 examination scores. The score in the "symptomatology/clinical reasoning" category of the GM-ITE also highly correlated with the PLAB 1 examination score. In addition, the scores in the "physical examination/ procedure," "medical interview/professionalism," and "disease knowledge" categories of the GM-ITE had modest correlation with the total score of PLAB 1 examination. The GM-ITE had better discriminative power than the PLAB 1 examination.

The GM-ITE is a newly developed residency in-training examination in Japan, which has been recognized as an important tool to assess residents' clinical knowledge acquisition. The results of the examination can also be used to improve clinical training programs at individual medical institutions. The questions are based on the content of the objectives of the clinical training presented by

Table 1 Characteristics of the Participants

Characteristics	Total N = 91	
	n	%
PGY		
PGY1	32	35.2%
PGY2	59	64.8%
Sex		
Male	22	24.2%
Female	69	75.8%
Hospital type		
Community	91	100.0%
University	0	0.0%
University branch	0	0.0%
Self-study time per day		
None	5	5.5%
0–30 min	36	39.6%
30–60 min	35	38.5%
60–90 min	10	11.0%
≥90 min	2	2.2%
ED duty per month		
0	2	2.2%
1–2	7	7.7%
3–5	68	74.7%
≥6	11	12.1%
Unknown	3	3.3%
Number of inpatients in charge		
0–4	10	11.0%
5–9	68	74.7%
10–14	8	8.8%
≥15	0	0.0%
Unknown	5	5.5%
Experience of taking international examinations		
USMLE	4	4.4%
PLAB	1	1.1%
Other examinations	0	0.0%
None	83	91.2%
Examination score	Mean	SD
GM-ITE score	30.2	5.5
PLAB score	119.5	11.7

Abbreviations: PGY, postgraduate year; ED, emergency department; PLAB, Professional and Linguistic Assessments Board; USMLE, United States Medical Licensing Examination; GM-ITE, General Medicine In-training Examination.

the MHLW.[7] The examination consists of 60 MCQs and has four subcategories, namely clinical reasoning, physical examination/procedure, communication/professionalism, and disease knowledge. Many of the questions ask about management in a medical setting to assess important clinical knowledge developed during the residency training.

Moreover, video questions have been incorporated from 2018 to further evaluate the content of the examination to assess practical experience (see Appendix 2). The examination development process is shown in Table 3. The examination was mostly developed by two committees, the question development committee (QDC; 23 members) and the peer-review committee (PRC; five members). Both committees are composed of experienced physicians in various fields. In addition, the QDC consists of an expert in examination analysis. In April, the process started with the analysis of the previous year's examination by the QDC. The analysis calculates the pass rate, probability of selecting each option, and polynomial correlation coefficient for each question, and the results are utilized in the preparation of the following year's examination.[14] From May to June, the QDC determined the subject areas and diseases, and request each member to create new questions. After the question development, the questions were reviewed by peer reviewers. In September, the PRC selected 60 questions from among 100 questions and asked the QDC members to revise the questions. A pilot study by senior medical residents (PGY 3–5) was conducted in late October. On the basis of the results of the pilot study, additional corrections were made to the questions. The final version of the examination was completed in November, and the examination was implemented from January to February. This process is repeated in the development of the annual examination, and some questions are used for the following year's examination.

The high correlation between the GM-ITE score and the PLAB part 1 score support the evidence of extrapolation component of the GM-ITE. Although the examinations are broadly similar in terms of assessment of clinical knowledge acquisition, they differ in terms of examinees and the purpose of the examination. The GM-ITE is administered to Japanese physicians who have graduated from medical school and are in postgraduate clinical training for 2 years. It measures the achievement during the training. On the other hand, the PLAB examination is a certification examination that checks the medical knowledge and skills required for medical school graduates outside the UK to work as physicians in the UK.[9] The UK has a postgraduate clinical training period called the Foundation Programme after graduation from medical school, which is equivalent to 2 years of postgraduate clinical education in Japan.[15] Therefore, the Japanese medical examination equivalent to the PLAB1 examination is the National Medical
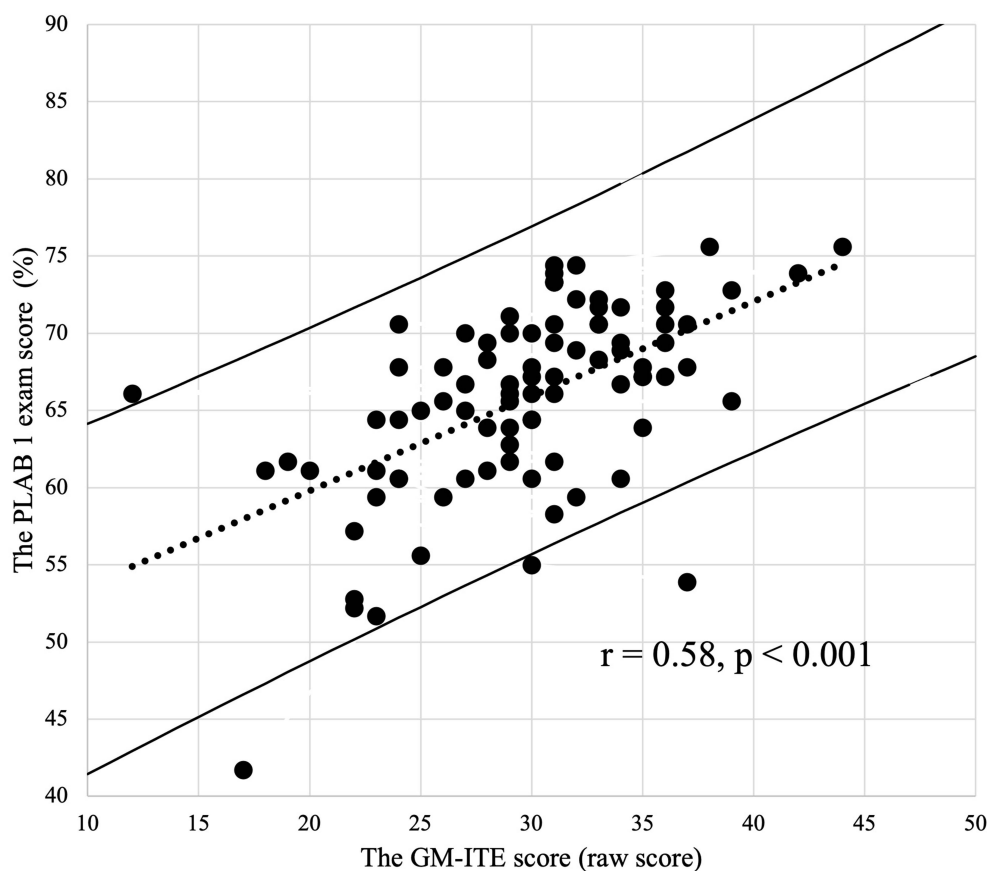
Figure 1 Scatter plot of the General Medicine In-training Examination (GM-ITE) and Professional and Linguistic Assessments Board (PLAB) 1 examination scores. The vertical axis is the PLAB 1 examination score, and the horizontal axis is the GM-ITE score.
Notes: The dotted line indicates the trend line. The two solid lines indicate the 95% prediction limits. r = Pearson correlation coefficient. $r \geq 0.40$ was considered satisfactory. Correlations were considered significant at the 0.01 level (two-tailed).
Abbreviations: GM-ITE, General Medicine In-training Examination; PLAB, Professional and Linguistic Assessments Board.

Practitioners Qualifying Examination (NMPQE), not the GM-ITE. In Japan, a knowledge-based examination is administered for medical licensing, which is the

Table 2 Pearson Correlation Coefficients of the General Medicine in-Training Examination and Professional and Linguistic Assessments Board 1 Examination Scores

GM-ITE	PLAB 1 Examination	
	r	p-value
Total	0.582	<0.001
Subcategories		
Medical interview/professionalism	0.252	0.015
Symptomatology/clinical reasoning	0.540	<0.001
Physical examination/procedure	0.376	<0.001
Disease knowledge	0.364	<0.001

Notes: r = Pearson correlation coefficient. $r \geq 0.40$ was considered satisfactory. Correlations were considered significant at the 0.01 level (two-tailed).
Abbreviations: PLAB, Professional and Linguistic Assessments Board; GM-ITE, General Medicine In-training Examination.

NMPQE. Medical students in Japan are required to pass the NMPQE to be registered as a physician. The NMPQE consists of MCQs that examine the essential clinical knowledge required during the 2-year mandatory training period. However, although the objectives and subjects of the two examinations differ, both share the same objective of assessing clinical diagnostic and management knowledge and real-world clinical applications. The results of this study support that the GM-ITE is well validated for assessing clinicians' clinical knowledge.

The correlation between the GM-ITE subcategory and PLAB 1 examination scores showed a high correlation only in the "symptomatology/clinical reasoning" category. The PLAB 1 examination asks about diagnosis, management, and real-world applications, and does not assess professionalism, communication, physical examination, and procedure.[16] This is the reason why the correlation is strong in clinical reasoning and not in medical interview/professionalism and physical examination/procedure. The PLAB 1
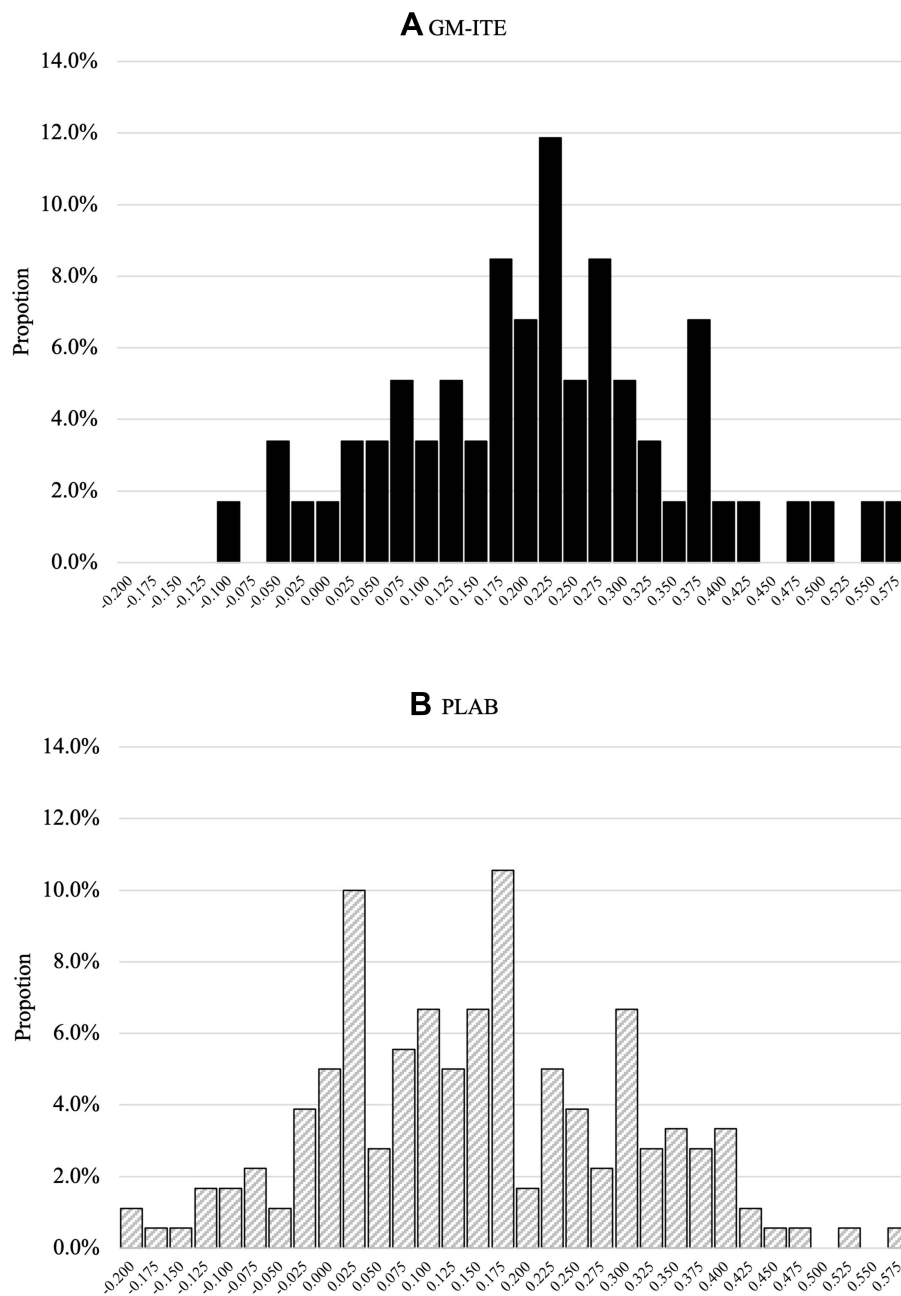
Figure 2 Discrimination indexes of the examination scores of the General Medicine In-training Examination (GM-ITE; **A**) and Professional and Linguistic Assessments Board 1 (PLAB; **B**). In both figures, the horizontal axis is the discrimination indices (DI), and the vertical axis is the proportion of questions within a certain DI range in the total questions. The figure with black bars shows the DI of GM-ITE, and the figure with shaded bars shows that of PLAB examination.
Notes: A discrimination index ≥ 0.20 indicates that the question has a high discriminatory power, and a discrimination index of >0.40 indicates that the question is a very good measure of the subject's qualifications.
Abbreviations: GM-ITE, General Medicine In-training Examination; PLAB, Professional and Linguistic Assessments Board.

examination, like the GM-ITE, assesses knowledge of many common diseases in various subcategories, not just internal and emergency medicine. Therefore, why the GM-ITE scores for the disease knowledge category did not correlate as strongly with the PLAB 1 examination scores is unclear.

The measurement of the discrimination index is an item analysis method that assesses the ability of individual questions to distinguish between high and low graders. Results show that both examinations have a high discrimination ability, but the GM-ITE has a sufficient discrimination index >0.20, which is higher than the PLAB 1 examination. The high discrimination index indicates that the GM-ITE, as an in-training examination, is an accurate indicator for assessing the development of clinical

Table 3 Development Process of the General Medicine in-Training Examination

April	Question Development Committee: -Examination analysis of the previous year's examination
May	Question Development Committee (pre-meeting): -Determination of the subject areas and diseases
Early June	Question Development Committee (phase 1): -Asks members to create new questions
Mid-June to late August	Question development period
Early September	Peer-review evaluation period

Mid-September	Peer-review committee (phase 1):
	   -Selection of 60 questions from among 100 questions
	   -Preparing revision plans

Late September	Peer-review committee (phase 2): -Asks committee members to revise questions for question development
Early October	Question development committee (phase 2): -Correction of the questions
Mid to late October	Pilot study by several senior residents

Early November	Question development committee (phase 3):
	   -Analysis of the result of the pilot study
	   -Minor correction of the questions

Mid-November	The final coordination
Late November	The completion of the question development process
Late January	Examination implementation

competence knowledge. In addition, inappropriate problems with negative discrimination indexes were also rare in the GM-ITE, indicating a high-quality test. However, the PLAB only determines whether examinees pass or fail and may not focus so much on discriminative performance.

Some limitations of the study should be noted. In this study, the PLAB 1 examination was translated into Japanese, but the equivalence of the pre- and post-translation examinations may not be preserved. Japanese translations are handled by an independent team, but no validation has been performed using back translation method.[17] Second, all the participants in this study were affiliated with community hospitals and did not include university residents. University residents account for approximately 40–50% of the residents in Japan, which weakens the generalizability of this study.[2] Third, this study recruited only small samples from all the examinees of the two examinations, and selection bias is possible.

Fourth, the participants in this study may have been well informed about the contents of the GM-ITE questions and may have scored higher in the GM-ITE than in the PLAB examination. Although the test questions are confidential and no past questions were used, because the GM-ITE is repeated by 87% of hospitals, examinees may have received prior information about the GM-ITE.

Conclusion

In conclusion, this study demonstrated the high correlation between the GM-ITE and PLAB part 1 examination, which provides incremental evidence of extrapolation in the GM-ITE. To further improve its validity, GM-ITE scores can be used to improve the qualities of training programs and policy-making on resident education in Japan. In the future, assessment of the real-word of performance of the GM-ITE examinees will further enhance the validity of the GM-ITE.[18,19] Whether differences in GM-ITE scores may result in differences in real-world

clinical performance and scores in specialist board examinations must be assessed. The existence of a validated examination will lead to improvements in the quality of residency education and even health care in Japan.

Data Sharing Statement

The corresponding author will respond to inquiries on the data analyses.

Ethics Approval and Consent to Participate

This study was approved by the ethics review board of Juntendo University School of Medicine, Tokyo, Japan. All the participants gave informed consent under an opt-out agreement.

Acknowledgments

Author Contributions

KN: Interpretation, Writing–Original Draft. YN: Conception, Study design, Execution, Acquisition of data, Interpretation, Writing–Review and Editing. MN: Conception, Analysis, Writing–Review and Editing. TS: Execution, Acquisition of data. RK: Execution, Acquisition of data. TO: Execution, Acquisition of data. YY: Execution, Acquisition of data. RM: Acquisition of data. HK: Writing–Review and Editing, Interpretation. YT: Conception, Study design, Interpretation. All authors contributed to data analysis, drafting, or revising the article, have agreed on the journal to which the article will be submitted, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

Funding

Disclosure

YN received an honorarium from the JAMEP as the GM-ITE project. YT and TO were responsible for the JAMEP director. KN received an honorarium from the JAMEP as a reviewer of GM-ITE. HK received an honorarium from the JAMEP as a speaker of the JAMEP lecture. TS, YY, and RM received an honorarium from the JAMEP as an examination preparer of GM-ITE. The authors report no other conflicts of interest in this work.

References

1. Teo A. The current state of medical education in Japan: a system under reform. *Med Educ*. 2007;41:302–308. doi:10.1111/j.1365-2929.2007.02691.x
2. Kozu T. Medical education in Japan. *Acad Med*. 2006;81:1069–1075. doi:10.1097/01.ACM.0000246682.45610.dd
3. Garibaldi RA, Subhiyah R, Moore ME, Waxman H. The in-training examination in internal medicine: an analysis of resident performance over time. *Ann Intern Med*. 2002;137:505–510. doi:10.7326/0003-4819-137-6-200209170-00011
4. Kinoshita K, Tsugawa Y, Shimizu T, et al. Impact of inpatient case-load, emergency department duties, and online learning resource on general medicine in-training examination scores in Japan. *Int J Gen Med*. 2015;8:355–360. doi:10.2147/IJGM.S81920
5. Nishizaki Y, Shimizu T, Shinozaki T, et al. Impact of general medicine rotation training on the in-training examination scores of 11, 244 Japanese resident physicians: a nationwide multi-center cross-sectional study. *BMC Med Educ*. 2020;20:426. doi:10.1186/s12909-020-02334-8
6. Hamstra SJ, Cuddy MM, Jurich D, et al. Exploring the association between USMLE scores and ACGME milestone ratings: a validity study using national data from emergency medicine. *Acad Med*. 2021;96(9):1324–1331. doi:10.1097/ACM.0000000000004207
7. Ministry of Health, Labour and Welfare. Objectives of clinical training. [Internet]. (in Japanese). Available from: https://www.mhlw.go.jp/topics/bukyoku/isei/rinsyo/keii/030818/030818b.html. Accessed December 5, 2020.
8. David BS, Richard EH. Reliability of scores and validity of score interpretations on written assessements. In: Holmboe ES, Hawkins RE, editors. *Practical Guide to the Evaluation of Clinical Competence*. 2nd ed. Philadelphia, PA: Mosby Elsevier; 2018:118–121.
9. Tiffin PA, Illing J, Kasim AS, McLachlan JC. Annual Review of Competence Progression (ARCP) performance of doctors who passed Professional and Linguistic Assessments Board (PLAB) tests compared with UK medical graduates: national data linkage study. *BMJ*. 2014;348:g2622. doi:10.1136/bmj.g2622
10. Wang H, Nugent R, Nugent C, Nugent K, Phy M. A commentary on the use of the internal medicine in-training examination. *APM*. 2009;122:879–883.
11. Muroya S, Ohde S, Takahashi O, Jacobs J, Fukui T. Differences in clinical knowledge levels between residents in two postgraduate rotation programmes in Japan. *BMC Med Educ*. 2021;21(1):226. doi:10.1186/s12909-021-02651-6

12. Shete A, Kausar A, Lakhkar K, Khan S. Item analysis: an evaluation of multiple choice questions in physiology examination. *J Contemp Med Edu*. 2015;3:106–109. doi:10.5455/jcme.201510 11041414

13. Chiavaroli N, Familari M. When majority doesn't rule: the use of discrimination indices to improve the quality of MCQs. *Biosci Educ*. 2016;17:1–7. doi:10.3108/beej.17.8

14. Mizuno A, Tsugawa Y, Shimizu T, et al. The impact of the hospital volume on the performance of residents on the general medicine in-training examination: a multicenter study in Japan. *Intern Med*. 2016;55:1553–1558. doi:10.2169/internalmedicine.55.6293

15. UK foundation programme. [Internet]. (in Japanese). Available from: https://foundationprogramme.nhs.uk. Accessed December 5, 2020.

16. General Medicine Council. PLAB test blueprint. [Internet]. (in Japanese). Available from: https://www.gmc-uk.org/registration-and-licensing/join-The-register/plab/plab-test-blueprint. Accessed December 5, 2020.

17. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. *Value Health*. 2005;8:94–104. doi:10.1111/j.1524-4733.2005.04054.x

18. Tiffin PA, Paton LW, Mwandigha LM, McLachlan JC, Illing J. Predicting fitness to practise events in international medical graduates who registered as UK doctors via the Professional and Linguistic Assessments Board (PLAB) system: a national cohort study. *BMC Med*. 2017;15:1–15. doi:10.1186/s12916-017-0829-1

19. Babbott SF, Beasley BW, Hinchey KT, Blotzer JW, Holmboe ES. The predictive validity of the internal medicine in-training examination. *Am J Med*. 2007;120:735–740. doi:10.1016/j.amjmed.2007.05.003