Virtual Patient Simulation in Junior-Level Japanese Resident Physicians: Attitudinal Questionnaire and Clinical Reasoning Enhancement

Brian Sanjay Heist*1,2 岸田 直樹*3
Gautam Deshpande*4 滨口 杉大*5 稲葉 竹俊*6
副島 久美子*7 小林 裕幸*8

Abstract
Background: Virtual Patient Simulation (VPS) is commonplace in Western medical education and, globally, increasingly
popular as a means to teaching clinical reasoning. VPS is not popular in Japan. This pilot study evaluates the receptiveness of Japanese junior resident physicians to VPS and the effectiveness of VPS on their clinical reasoning.

Methods: Each first-year junior resident physician at 4 residency programs (n=54) was assigned a randomized sequence to complete 5 VPS modules. In accordance, online access was provided to each subsequent module at 6 day intervals. We evaluated participants’ change in performance on pre-post administration of a validated instrument of diagnostic reasoning. We developed a scoring rubric and evaluated change in efficiency of selecting items in the history and physical examination portions of the modules. Upon completing the module series, participants received a validated questionnaire on the effectiveness of VPS.

Results: Pre-posttest improvement in diagnostic reasoning scores was not statistically significant (p=0.07). Improvement in efficiency from 1st to 5th modules was significant for history taking (p<0.01), but not physical examination (p=0.697). Mean perceived effectiveness of VPS was superior to traditional methods (mean 3.58-3.71 on 1-5 Likert scale) for understanding methods of diagnosis, medical management, generation of differential diagnosis, and knowledge about clinical cases.

Conclusions: First-year resident physicians demonstrated enthusiasm for VPS. Improvement on the validated measure of diagnostic reasoning was not confirmed and thus, it is premature to draw conclusions on the effectiveness of VPS amongst Japanese junior resident physicians. Larger scale follow-up investigation is warranted.

Key words: Virtual Patient Simulation, Clinical Reasoning, Junior Residency Education

Background

Japanese medical education has historically developed expertise in fundamental research and highly specialized clinical skills(1). However, it has been less effective in developing general clinical skills, as evidenced by a recent national questionnaire that Tokuda and colleagues administered to all PGY1 Japanese residents. “Of the 2429 respondents to the questionnaire, 406 (17%) agreed or strongly agreed that they were confident that they had acquired the general clinical skills required to begin postgraduate training. Overall, confidence in preparedness for postgraduate training in other clinical areas was indicated by 686 respondents (29%) for basic knowledge of diagnosis and management of common conditions, [and] 465 (19%) for skills for applying evidence-based medicine to clinical care . . . .”(2) Consistent with the questionnaire results, Japanese junior-level residents have embraced publications that attempt to teach physician trainees how to approach common (and sometimes uncommon) clinical scenarios. Examples include “月刊誌レジデントノート, 誰も教えてくれなかった診断学”, “Dr 宮城の教育回診実況中継—ホンモノの診察技法と疾患を劇的に絞り込む思考プロセス”, and “ティアニー先生の診断入門” In many instances, the authors have been influenced directly or indirectly by Western training and consistently they attempt to teach not just diagnostic algorithms, but skills of clinical reasoning (CR).

The existence of such texts focused on teaching CR contrasts the Western medical education landscape. In the West, CR has historically been taught through close feedback on written or oral presentations and through regularly held case-based discussions(3). Though neither method has undergone formal validation studies(4), each prioritizes education of hypothetico-deductive CR essential early in clinical training and is widely presumed to teach CR(3). At the same time, in the West, over the past 4 decades computer-based Virtual Patient Simulation (VPS) has increasingly supplemented tradition(5). VPS has been defined as “computer programs that simulate real-life clinical scenarios in which the learner acts as healthcare professional obtaining a history and a physical exam and making diagnostic and therapeutic decisions”(6). More recently, researchers have begun to evaluate the roles for e-learning within medical education and recognized that perhaps
the best contribution is VPS’s teaching of CR\textsuperscript{4,7}. Beyond Europe and North America, nations such as Korea are developing virtual patient module repositories\textsuperscript{9}, and international collaborations are constructing VPS for developing nations with limited clinical teaching resources\textsuperscript{9}. In contrast, Japanese medical teaching institutions thus far have not embraced the use of VPS. The basis is unclear. VPS popularity and effectiveness amongst Japanese physicians-in-training is anticipated for several reasons. Japanese trainees are technologically savvy with access to high-speed internet. Japanese trainees are receptive to CR teaching methods adopted from the West as evidenced by the publications referenced above and the popularity of a case discussion teaching format at several Japanese residency programs. There is great variation in the training experience of Japanese physicians\textsuperscript{10, 11}. Finally, Japanese residents spend less than 7% of their time engaged in educational activities\textsuperscript{11}. Recognition of these issues during the primary investigator’s tenure at Teine Keijinkai Hospital served as inspiration for this study.

Presently, VPS most commonly targets medical students during their clinical years, but modules targeting all levels of physicians exist. VPS for post-graduate trainees has received interest from both developers and users\textsuperscript{7}, and been recognized as an area requiring further research. Additionally, to our knowledge only 6 published VPS intervention studies have evaluated CR. Methodologies and outcomes measures vary, and conclusive evidence of CR enhancement is lacking\textsuperscript{5, 12, 13}. Measures of clinical reasoning have included scoring of the user’s data collection and decision-making during the simulation\textsuperscript{14-18}, and external testing limited to standardized patients and written examinations on the specific clinical scenarios addressed by the VPS intervention\textsuperscript{12, 19-21}.

The objectives of this study were two-fold. First, we sought to assess the receptiveness of Japanese physicians-in-training to VPS. Second, we sought to measure the effect of VPS on participants’ diagnostic reasoning. We were more interested in their fundamental reasoning skills and thought processes than proficiency with specific clinical scenarios. In turn, we used

1) a scoring system for the history and physical examinations performed during the VPS modules and

2) pre-post evaluation with a previously validated CR assessment instrument not specific to the clinical scenarios addressed.

**Methods**

We conducted a randomized trial on change in diagnostic reasoning with a VPS system developed for this study. This was accompanied by a questionnaire regarding the VPS experience. The study was conducted from October 2012 – May 2013. Institutional Review Boards at the University of Pittsburgh and St. Luke’s International Hospital, and comparable governing boards at Teine Keijinkai Hospital, Mito Kyodo General Hospital, and Eibetsu City Hospital approved the study.

**Virtual Patient Simulation**

Five VPS modules were developed by the Computer Science Department at the Tokyo University of Technology using Adobe Flash\textsuperscript{©} software. Each module was a computerized transformation of a clinical case presented at a PGY1 Western style case discussion at Teine Keijinkai Hospital from 2007 – 2010, and prioritized the elements in **Table 1**. Prioritized criteria, such as the trainee’s misunderstanding of a concept or sense he/she could have managed a situation better (see appendix A for a complete list), are consistent with literature on trainees’ top principles of virtual patient design\textsuperscript{22}, and the recognition that native
Table 1: Paired Wilcoxon Signed-Rank Test for DTI Performance

<table>
<thead>
<tr>
<th></th>
<th>Pre-VPS score</th>
<th>Post-VPS score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>148.62</td>
<td>156.86</td>
</tr>
<tr>
<td>Median</td>
<td>150</td>
<td>161</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>18.89</td>
<td>14.85</td>
</tr>
</tbody>
</table>

P-value = 0.07

language case presentations using scenarios and management reflective of the local culture are more effective than imported clinical scenarios.

Each module consisted of a brief history of present illness, followed by interactive lists of history-taking and physical examination items, a case summarization exercise, and then a linear pathway of interactive slides addressing further evaluation and management. All sections reflected actual evaluation and management performed. Instructions accompanied the respective sections of the module, and immediate feedback was provided on the appropriateness of each item selected using symbols to reflect essential, relevant, reasonable, unnecessary, and inappropriate choices, respectively. Teaching points were embedded in the management section and concluding slides of each module. To further encourage hypothetico-deductive reasoning, participants were asked to enter lists of differential diagnoses early in the diagnostic portion. A reference bank of symptoms, risk factors, and examination findings for relevant clinical conditions was similarly provided to focus the module on the thought process of diagnosis and enable selection of appropriate items by users without detailed knowledge of the topic.

In consideration of the competing responsibilities of Japanese junior resident physicians, VPS modules for this study were designed with the goal that a user already knowledgeable on a given topic could complete the mandatory module components within ten minutes. Module content was completely in Japanese. Access to the modules is available by contacting the PI.

Participants

In September - October 2012, All PGY1 resident physicians (n=54) rotating at 2 institutions in Hokkaido, Teine Keijinkai Hospital (15 residents) and Ebetsu City Hospital (5 residents), and 2 institutions in the Kanto region, St. Luke’s International Hospital (24 residents) and Mito Kyodo General Hospital (10 residents) were invited by local site coordinators to participate in the study. Informed consent was obtained online before commencing study participation. Each participant was compensated 1000 yen for each module completed.

Study Design

Local site coordinators verbally explained the study to each potential participant and transferred email addresses for all potential participants to the data management server administrator at the Computer Science Department at the Tokyo University of Technology (further referred to as ‘server administrator’). Within two weeks, the server administrator emailed each potential participant with further explanation of the study.

At the start of the study, the server administrator emailed each potential participant a website address containing the informed consent confirmation and then the pre-test DTI (the Diagnostic Thinking Inventory is described below). Upon completion of the DTI, the participant received access to download the first VPS module. Each potential participant was assigned a randomized sequence of the 5 modules as generated by the University of Pittsburgh Department of Biostatistics. Download site addresses for subsequent assigned modules were emailed at 6 day intervals and participants were requested to complete each
module before receiving access to the next module. This interval was selected in anticipation of resident physician behavior in the non-study setting and to enable completion of all modules within one month. For security reasons, at St. Luke’s International Hospital, modules were delivered to each participant via USB memory stick. Upon completion of the final module, participants were asked to again complete the DTI, and then an online questionnaire about the VPS experience. All responses entered into the computer were automatically transferred to the data management server at the completion of each DTI, VPS module, and questionnaire, and labeled with participant ID number.

This study has no control arm of participants not assigned to VPS modules. Given the lack of prior data on VPS in Japanese physician trainees, a power calculation could not accurately be performed to estimate a suitable population size for such a trial. In turn, this study was conceived as a pilot study.

Assessment of the Intervention Questionnaire

Upon completion of the final module, each study participant received an online questionnaire about the VPS experience. Likert scale questions comparing the VPS experience to traditional learning methods were based on a questionnaire validated amongst American medical students completing VPS modules during their pediatrics rotation. Two investigators (BH, HK) performed the Japanese translation with slight adaptation to the current study population. Additionally, participants were queried about the time required for module completion, usability, difficulty, and applicability of the modules. The questionnaire can be accessed at: http://aruefluid.sakura.ne.jp/clinicalCaseModule/clinicalCaseAncate/index.html.

Measures of Diagnostic Reasoning

Two instruments were employed in this study: the Diagnostic Thinking Inventory and Item Selection in History-taking and Physical Examination. Clinical reasoning is a term commonly used, but at risk of varied interpretation. To help reconcile our assessment methodologies and avoid misunderstanding amongst readers, we refer to The Thinker’s Guide to Clinical Reasoning. “Clinical reasoning can be defined as thinking through the various aspects of patient care to arrive at a reasonable decision regarding the prevention, diagnosis, or treatment of a clinical problem in a specific patient.”

Diagnostic Thinking Inventory (DTI): The DTI is a validated test of 41 items that probe a physician’s reasoning during the diagnostic process. Each test item presents a brief clinically relevant stem followed by 2 alternatives that reflect ‘low knowledge’ (poorly effective organization of knowledge) and ‘high knowledge’ (effectively elaborated and compiled knowledge), respectively. The test-taker selects from six options where his/her decision-making lies between the 2 alternatives, and receives a score from 1 to 6. Left-right orientation of the superior and inferior alternatives varies by item. Appendix B contains a sample test item.

<table>
<thead>
<tr>
<th>Appendix A: Elements of Western style Case discussion of clinical scenario prioritized in VPS module design</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case selection criteria for Case Discussion:</strong></td>
</tr>
<tr>
<td>Resident misunderstood a concept</td>
</tr>
<tr>
<td>Resident didn’t know what to do</td>
</tr>
<tr>
<td>Resident thought he/she could have done better</td>
</tr>
<tr>
<td>Resident thought the differential diagnosis was interesting</td>
</tr>
<tr>
<td>Resident expected fellow residents to encounter the same situation</td>
</tr>
<tr>
<td>Additionally,</td>
</tr>
<tr>
<td>*Ideally, details of the HPI, PMH, Meds, etc. were known</td>
</tr>
</tbody>
</table>


伝統的な教材に比べ、e ラーニングモジュールを用いた方法は、症例の理解を深めるために、どの程度……
有効ですか？

どのような流れで症例の診療を行うかについて理解を深めるために、どの程度……
有効ですか？

臨床症例に対する鑑別診断を挙げる能力を高めるために、どの程度有効ですか？

臨床症例に関わる知識を得るために、どの程度有効ですか？

問診技能を高めるために、どの程度……
有効ですか？

身体診察技能を高めるために、どの程度……
有効ですか？

優れた臨床技能（診療）の手本として……
有効ですか？

mean rating (SD)

3.71 (0.98)

3.63 (1.21)

3.67 (0.96)

3.58 (1.02)

3.21 (1.01)

2.46 (0.83)

3.42 (1.14)

Figure 2. Post-VPS Questionnaire : Effectiveness of VPS modules to Traditional Educational Methods (n=24)
Respondents used a 5 point scale: 1 = 全く効果がない 3 = 同じ程度効果がある 5 = とても効果がある

*The case was NOT too rare or with a distracting long preceding hospital course

Focal points of Discussion
practicing the thought process of hypothetical-deductive clinical reasoning.
case summarization
teaching points including “cannot miss” differential diagnoses, and important physical exam and laboratory data clues

The DTI is especially suited for less experienced clinicians, such as those in our study, whose domain specific clinical knowledge base may confound interpretation of reasoning skills observed during specific clinical scenarios. The DTI has been utilized as an outcome measure of CR in numerous studies, including an ongoing study of VPS(27).

A Japanese version of the DTI was prepared for this study. Two investigators (BH, HK)

Appendix B: Sample question from Diagnostic Thinking Inventory.
After reading the question stem, from the 6 boxes, the physician selects the box where his/her thinking lies on the spectrum between the two alternatives.

患者に症状を説明してもらう際。
When having a patient explain his/her symptoms,

I think about the symptoms using the words that the patient actually used

The patient’s own expressions with abstract terminology (for example “acute” for “duration of 4 days”, “bilateral” for “both hands/two hands”)
together performed the forward translation and a third physician bilingual in Japanese and English performed the back translation. The original author of the DTI, Georges Bordage, confirmed the accuracy of the translation.

**Item selection in History-Taking and Physical Examination:** Hypothetico-deductive reasoning involves developing differential diagnoses early in patient evaluation and then asking questions and performing examinations to support or refute these diagnoses. By consistently practicing this process with feedback on the quality of their questions and examination choices, less experienced physician trainees can develop rich illness scripts of addressed clinical conditions. Correct illness scripts are essential to correct application of pattern recognition, the clinical reasoning method used most commonly by experienced clinicians. The Western style case discussion, which forms the basis for this VPS design, is foremost an exercise in hypothetico-deductive reasoning. Importantly, the facilitating physician guides participants through the case diagnosis, but also evaluates the participants' CR in part through the close-ended questions they ask about symptoms, enabling conditions (i.e. risk factors), physical examination findings, and so forth. Similarly, in this study as this VPS guided the hypothetico-deductive diagnostic evaluation of each clinical case, it assessed the user’s reasoning through application of a history and physical examination item scoring system developed for this study. Consistent with displayed feedback symbols described above, each item was assigned a score from (−) 2 points for inappropriate items to (+) 2 points for items essential to diagnosis (e.g. inquiry about medications in an elderly diabetic patient found down and poorly responsive).

Using reference texts on clinical features of the addressed scenarios, including the JAMA publication, “論理的診察技術” where applicable, as well as personal clinical experience, two investigators (BH, NK) independently applied scores to each item selected in the VPS modules and then negotiated any disagreements. Scores were then reviewed by other research team members. Items selected by participants and corresponding scores were automatically recorded and summed by the VPS system, enabling comparison of performance on different modules.

Comparable outcome measures have been employed previously in VPS based clinical reasoning assessment. In our study, all five involved research team members are experienced faculty clinician educators with personal experience facilitating case discussions. In turn, while this evaluation method has not undergone prior validation studies, it was felt to have high content validity. Results consistent with those from the questionnaire and DTI, the

**Figure 1:** Flow of participants through study. Due to loss of data during electronic transmission to server, n reflects confirmed participation only. Completion of Questionnaire prior to Post-Test Diagnostic Thinking Inventory was not possible.

<table>
<thead>
<tr>
<th>Introduction to Study (n=54)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-VPS Diagnostic Thinking Inventory (n=40)</td>
</tr>
<tr>
<td>VPS module #1 (n=39)</td>
</tr>
<tr>
<td>VPS modules #2-4</td>
</tr>
<tr>
<td>VPS module #5 (n=29)</td>
</tr>
<tr>
<td>Post-VPS Diagnostic Thinking Inventory (n=13)</td>
</tr>
<tr>
<td>Questionnaire (n=24)</td>
</tr>
</tbody>
</table>
content of which addresses hypothetico-
deductive reasoning, would contribute
concurrent validity to this methodology\(^\text{30}\).

### Statistical Analysis

Non-parametric tests were conducted in
consideration of the sample size. Mean and
standard deviation were computed for Likert
scale items on the post–VPS questionnaire. For
DTI results, the Paired Wilcoxon Signed–Rank
Test was conducted to compare pre–VPS and
post–VPS scores. For scores tabulated from
items selected in history-taking and physical
examination sections of the modules, the Non-
paired Wilcoxon Rank Sum Test was conducted
to compare collective performance on module #1
and module #5. A paired test was not used
because of unavoidable variation in difficulty
between VPS modules due to differing clinical
scenarios. By comparing performance of the
participants collectively on the respective
modules, it was anticipated that such variation
would average out. All tests were 2–sided and
the level of statistical significance was specified
at 0.05.

**Results**

Of the 54 invited participants, completion of the
pre–VPS DTI was confirmed for 40 participants,
module 1 for 39 participants, module 5 for 24
participants, post–VPS DTI for 13 patients, and
the questionnaire for 24 participants (Figure 1). Completion of questionnaire prior to the DTI
was expected to be impossible. Upon
commencing results analysis in early February
2013, two weeks after completion of the planned
study period, it was learned that some data had
been lost in transmission from participant
computers to the data management server and
that some participants had not received emails
from the server. These issues were believed to
result from server malfunction and poor
compatibility between the server and some
participant email accounts (e.g. Gmail, Yahoo,
Hotmail). Relevant participants were identified
by the server administrator and contacted by
local site coordinators, resulting in extension of
module access until early May 2013. Participants did not repeat any portions of the
study. Data for 2 participants were added due to

**Figure 3.** Post–VPS Questionnaire: Overall Impression, Usability, Difficulty, Deepening of Clinical Reasoning, and Applicability of VPS Modules (\(n = 24\))
"VPS in Japanese Residents" 「初期研修医に対する VPS」

Table 2:

<table>
<thead>
<tr>
<th>Section Scored</th>
<th>All Module 1 (n = 39) vs All module 5 (n = 29)</th>
<th>Matching Module 1 (n = 29) vs Module 5 (n = 29)</th>
<th>Balanced Module 1 (n = 22) vs Module 5 (n = 29)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean (SD)</td>
<td>mean (SD)</td>
<td>mean (SD)</td>
</tr>
<tr>
<td></td>
<td>Module 1</td>
<td>Module 5</td>
<td>p-value</td>
</tr>
<tr>
<td>History taking</td>
<td>0.598</td>
<td>0.718</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>(0.207)</td>
<td>(0.203)</td>
<td></td>
</tr>
<tr>
<td>Physical examination</td>
<td>0.388</td>
<td>0.517</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>(0.697)</td>
<td>(0.171)</td>
<td></td>
</tr>
<tr>
<td>History taking +</td>
<td>0.479</td>
<td>0.594</td>
<td>0.23</td>
</tr>
<tr>
<td>Physical exam</td>
<td>(0.393)</td>
<td>(0.393)</td>
<td></td>
</tr>
</tbody>
</table>

this modification.

Questionnaire

Nine of 24 respondents previously used e-learning and 10/24 respondents previously studied clinical reasoning. Questions in Figure 3 were adapted from a validated questionnaire of American 3rd year medical students completing a VPS module series24. Overall, the respondents reported this VPS system was superior to traditional learning methodologies for developing skills related to CR and building clinical knowledge. The VPS modules were perceived inferior for learning physical examination skills and slightly superior for learning history taking. All results are approximately 0.20 points lower than those for the above referenced study of American medical students24. A median duration of 10–15 minutes was required for module completion, though 3:24 respondents (13%) required over 20 minutes. Nineteen of 24 respondents (79%) reported module completion was an effective use of time.

Figure 3 contains questions addressing overall impressions and applicability of the VPS module content. Most respondents reported a positive impression of the modules (88%), and found them easy to use (75%), and challenging (71%). Seventy-four and 75 percent of respondents, respectively, felt the modules deepened their understanding of CR and that the CR was applicable to clinical practice.

Change in Diagnostic Reasoning

Pre-VPS and post-VPS DTI results were recorded for 13 participants and compared using the Paired Wilcoxon Signed–Rank Test (Table 1). Statistically non-significant improvement in DTI score (p 0.07) was observed. As described above, it is believed that additional results were lost in transmission between participant computers and the data management server.

Table 2 compares the participants’ selection of history taking and physical examination items on the first and fifth VPS modules. Due to incomplete participation, the distribution of the 5 clinical cases completed for Module 1 did not match those completed for module 5. Additionally, completion of module 1 was confirmed for 39 participants, but completion of module 5 was confirmed for only 29 participants.

To address the potential influence of these discrepancies, the Wilcoxon Rank Sum Test was performed 3 times: 1) all 39 completed module 1 vs all 29 completed module 5, 2) module 1 vs module 5 for the 29 participants completing both modules, 3) module 1 vs module 5 for 22 participants, after eliminating data from 7 participants (through sequential review of assigned ID numbers) to improve the balance of
clinical scenarios. In all three cases, statistically significant improvement from module 1 to module 5 was observed for history taking items, but not for physical examination items, or the combination of both history taking and physical examination items. Notably, there was wide variation in the selection of physical examination items, and in the free texted questionnaire comments, some participants felt that it should be appropriate to perform certain physical exams even when not pertinent to diagnosis of the patient’s presenting condition (e.g. lung auscultation in previously healthy 16 y/o boy with appendicitis).

Discussion

Globally, substantial research has demonstrated the effectiveness of virtual patient simulation in medical education\textsuperscript{31}. Reasons underlying the lack of this modality in Japanese medical education are unclear. In our small scale pilot study, Japanese PGY1 level resident physicians responded positively to VPS, but failed to demonstrate statistically significant improvement in diagnostic reasoning after completion of the VPS modules.

Regarding the post-VPS questionnaire, participants reported the modules superior to traditional educational materials for developing CR and building clinical knowledge. Mean scores were consistently approximately 0.2 / 5 Likert scale points less than scores from American medical students completing VPS modules while rotating in pediatrics. The similarity is interesting when considering that the American study involved medical students with slightly different motivations (e.g. relatively little responsibility for patient care, end-of-rotation examination) than our participants, modules requiring over twice as much time (approximately 30-50 minutes compared to 10-15 minute median duration in our study), clinical scenario topics matching the rotation, and protected time for module completion\textsuperscript{32}. When considering Japanese junior residents’ busy schedules and competing responsibilities, these data, together with the positive feedback on usability and clinical applicability of the modules, are encouraging. We anticipate further improved response if our modules were integrated into corresponding rotations. We also anticipate enthusiasm amongst Japanese medical students, though the contrasting curricular structure from Western medical schools, including many months of study dedicated to the national examination at the conclusion of Japanese medical school, complicates the timing for integration of VPS.

The participants’ feedback that the modules are effective in building clinical knowledge is consistent with prior research showing VPS superiority to lectures for this purpose\textsuperscript{12}. Additionally, most questionnaire respondents felt the VPS deepened their understanding of CR and that the CR was applicable to clinical practice. This is relevant since medical education literature suggests trainees’ perception of their learning is a valid surrogate for testing in computer assisted instruction\textsuperscript{33}.

Our study employed two measures of diagnostic reasoning, the DTI, a previously validated test of CR, and an automated scoring system we developed to measure the efficiency of selecting appropriate history taking and physical examination items within the modules. DTI data analysis was limited to 13 participants due to data loss, and the difference in paired pre-test / post-test performance failed to achieve statistical significance (\(p > 0.07\)). Our automated scoring system has not undergone prior validation studies, but was felt to have high content validity as explained above. Data from the modules’ automated scoring system was analyzed for over twice as many participants.
Statistically significant improvement was observed in history taking, but not physical examination. These findings were maintained after modification of the data set to decrease the impact of variation in module difficulty. As evidenced by the relatively large S.D., participants’ selection of physical examination items varied widely. In clinical practice, elements of the physical examination such as cardiopulmonary auscultation are performed partly out of routine and to appease patient’s expectations, even when there is no impact on CR. In the questionnaire, multiple participants complained about scoring penalties for such exams, indicating an area for improvement in future versions of our VPS modules and also a potential reason underlying participants’ lack of improvement in the physical examination portion of the modules.

Our results have potential implications for the improvement of Japanese medical education. As evidenced by the popularity of written texts teaching clinical medicine including clinical knowledge and diagnostic reasoning, Japanese physicians-in-training are highly motivated to improve their diagnostic skills. Consistent with endeavors in other countries, particularly in the setting of limited clinical teaching resources and varied training experience across residency programs, VPS offers an efficient means to supplement the clinical training of physicians across Japan. In the case of the VPS modules used herein, the core elements of the Western style case discussion were condensed into a median 10–15 minute learning experience, feasible within the busy schedule of Japanese Junior Residents. While this pilot scale study failed to demonstrate statistically significant improvement on a validated measure of CR, the participants’ enthusiasm for this modality and the improved efficiency in the history taking component of hypothetic–deductive diagnosis provide motivation for future larger scale investigation.

This study has several important limitations. Firstly, the multi-institutional sample size was felt to be sufficient for this pilot scale study, but there were complications with email correspondence between server administrator and participants, and data transfer to the server resulting in data loss and extension of trial duration. The delayed completion of modules by some participants is important since CR and attitude toward VPS change with increased clinical experience. Secondly, there was no control arm of participants who did not complete VPS modules, so it is possible that the observed improved diagnostic reasoning did not result from the intervention. Thirdly, the results may be influenced by participant selection bias since the participating residency programs are reputed for their emphasis on clinical teaching. In turn, though participants were in their PGY1 year and graduated from medical schools across Japan, the findings here may not be generalizable to other PGY1 physicians.

Conclusions

Our small scale pilot study demonstrates enthusiasm of Japanese PGY1 physicians for virtual patient simulation as a method to improve diagnostic reasoning, but failed to confirm improvement in their diagnostic reasoning after completing a series of VPS modules. The next steps in this endeavor include refinement of the VPS module design in response to participant feedback and execution of a larger scale study with a control arm.

Funding

This study was funded jointly by St. Luke’s Life Science Institute and Teine Keijinkai Hospital.
Acknowledgements

We would like to express our gratitude to Dr. Haruka Matsubara Torok for performing the back translation of the Diagnostic Thinking Inventory.

References

23. Fors UG, Muntean V, Botezatu M, Zary N.


29) 竹本英 訳. 理論的診察の技術ー エビデンスに基づく診断のノウハウ. 日経 BP 社; JAMA; 2010.


