

# Assessment of Clinical Reasoning by Listening to Case Presentations: VSOP Method for Better Feedback

Hiroataka Onishi

International Research Center for Medical Education, Graduate School of Medicine, University of Tokyo, Tokyo, Japan.

**ABSTRACT:** Case presentation is used as a teaching and learning tool in almost all clinical education, and it is also associated with clinical reasoning ability. Despite this, no specific assessment tool utilizing case presentations has yet been established. SNAPPS (summarize, narrow, analyze, probe, plan, and select) and the One-minute Preceptor are well-known educational tools for teaching how to improve consultations. However, these tools do not include a specific rating scale to determine the diagnostic reasoning level. Mini clinical evaluation exercise (Mini-CEX) and RIME (reporter, interpreter, manager, and educator) are comprehensive assessment tools with appropriate reliability and validity. The vague, structured, organized and pertinent (VSOP) model, previously proposed in Japan and derived from RIME model, is a tool for formative assessment and teaching of trainees through case presentations. Uses of the VSOP model in real settings are also discussed.

**KEYWORDS:** work-based assessment, outpatient education, formative assessment, reliability, validity, global rating scale

**CITATION:** Onishi. Assessment of Clinical Reasoning by Listening to Case Presentations: VSOP Method for Better Feedback. *Journal of Medical Education and Curricular Development* 2016;3:125–131 doi:10.4137/JMECD.S30035.

**TYPE:** Original Research

**RECEIVED:** February 26, 2016. **RESUBMITTED:** August 21, 2016. **ACCEPTED FOR PUBLICATION:** August 23, 2016.

**ACADEMIC EDITOR:** Steven R. Myers, Editor in Chief

**PEER REVIEW:** Three peer reviewers contributed to the peer review report. Reviewers' reports totaled 865 words, excluding any confidential comments to the academic editor.

**FUNDING:** This work was supported by JSPS KAKENHI Grant Number 21590757. The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

**COMPETING INTERESTS:** Author discloses no potential conflicts of interest.

**COPYRIGHT:** © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC 3.0 License.

**CORRESPONDENCE:** onishi-hiroataka@umin.ac.jp

Paper subject to independent expert single-blind peer review. All editorial decisions made by independent academic editor. Upon submission manuscript was subject to anti-plagiarism scanning. Prior to publication all authors have given signed confirmation of agreement to article publication and compliance with all applicable ethical and legal requirements, including the accuracy of author and contributor information, disclosure of competing interests and funding sources, compliance with ethical requirements relating to human and animal study participants, and compliance with any copyright requirements of third parties. This journal is a member of the Committee on Publication Ethics (COPE).

Provenance: the author was invited to submit this paper.

Published by Libertas Academica. Learn more about this journal.

## Introduction

Case presentation is closely related to clinical reasoning ability. The focus of this paper is on diagnostic reasoning ability, the first half of the clinical reasoning process, and the key to the following phase of clinical decision-making. When a trainee prepares for a case presentation after a patient's initial presentation, he/she must sort out and organize information along with diagnostic hypotheses. Preparation for the case presentation works to facilitate reflection, which will promote diagnostic reasoning ability.<sup>1</sup>

In clinical settings where time constraints are an important consideration, teaching tools using a dialog format are well accepted. Two typical tools are the One-minute Preceptor<sup>2</sup> and summarize, narrow, analyze, probe, plan, and select (SNAPPS).<sup>3</sup> Both models are used by preceptors who listen to case presentations from trainees and give advice and suggestions for improvement. Diagnostic reasoning is one of the most important teaching points for both models. However, these models do not explicitly contain a component of assessment. Thus, it is thought that preceptors assess each case presentation empirically and formatively.

Key indicators to assess diagnostic reasoning ability are the diagnostic hypotheses (differential diagnoses) in a case presentation and the quantity and quality of information collected to articulate those diagnostic hypotheses. Elstein et al<sup>4</sup> showed that low performers in diagnostic reasoning tend

to collect thorough information. Gruppen et al<sup>5</sup> found that medical students who had the correct diagnosis in mind were four to nine times more likely to reach the correct diagnosis after history taking.

From the developmental viewpoint, it is suggested that the more diagnostic ability a trainee has, the better the quality of his/her case presentations will be. Initially, trainees may not remember which information they should gather nor how best to sort out information in the case presentation. After experiencing several cases, they develop an understanding of the series of information required in a case presentation for a trainer to understand the case and give advice. If a trainee is able to consider key differential diagnoses (KDDs) as hypotheses, those hypotheses will become the guide to ask closed questions known as pertinent positive/negative questions. When all pertinent positive and negative questions are asked and such information is well organized into the case presentation, the development of diagnostic reasoning ability is considered complete.

Thus, categorizing case presentations by levels of diagnostic reasoning ability would become the basis for using case presentations to assess clinical reasoning ability. However, different specialties or different trainers might have different criteria for assessing case presentations. Key points of case presentation may differ depending on which information the trainer believes to be important. Each trainer has her/his own

**Table 1.** Global rating scale for diagnostic reasoning.

LEVEL	CONDITION OF PRESENTATION	LEVELS OF THE PRESENTER	FEEDBACK FROM THE TRAINER TO THE PRESENTER
1	Lack of essential information of the case or inappropriate definition or reliability of information about S/S.	Lack of basic clinical skills for H&P or required information for case presentation.	Ask the presenter about inappropriate terms or essential information of case presentation. Offer one-to-one practice to the presenter.
2	Insufficient or unordered information.	Inability to capture each piece of information or organize information for the case.	Point out what is missing in case presentation. Ask the presenter to practice case presentation.
3	Essential information is well covered but KDDs are not well listed.	Able to report the case but unable to interpret the patient problems.	Give positive feedback for complete information. Ask the presenter to summarize the presentation and KDDs.
4	KDDs are covered but pertinent positive and negative S/S are insufficient.	During H&P no relevant S/S to KDDs obtained.	Give positive feedback for KDDs and specific feedback for pertinent positive and negative S/S.
5	Pertinent positive and negative S/S relevant to KDDs are covered.	Through H&P whole picture of the case and its KDDs are clearly described.	Give positive feedback for a good presentation. Ask the presenter to specify the lesson learned from the case.

**Abbreviations:** H&P, history taking and physical examination; S/S, signs and symptoms; KDDs, key differential diagnoses.

KDDs in mind when assessing a trainee. A global rating scale (GRS) using the expertise in each specialty might work as a good tool for assessing case presentations to avoid level differences among different specialties.

An example of such a GRS is the RIME method proposed by Pangaro,<sup>6</sup> a generic assessment tool utilized in various rotation sites. RIME stands for reporter, interpreter, manager, and educator, easy-to-remember descriptors for each level of clinical trainees. It is important for the GRS for assessment of diagnostic reasoning to also have easy-to-remember descriptors that can be utilized in different rotation sites.

Another point is the importance of work-based assessment (WBA), the ultimate level in Miller's pyramid for clinical assessment.<sup>7</sup> Mini clinical evaluation exercise (mini-CEX), a WBA tool with aggregation of several GRSs, is one of the most popular WBAs.<sup>8</sup> Mini-CEX includes diagnostic reasoning ability as "clinical judgment," yet does not explain how such clinical judgment is assessed for each clinical encounter.

Onishi<sup>9</sup> developed a WBA tool for diagnostic reasoning by listening to case presentations mainly in outpatient settings. The original version was developed in Japanese and it was translated into English (Table 1). The aims of this article are: (1) to theorize this assessment tool, (2) to change the original 5-point scale model into a 4-point scale model, and (3) to discuss the direction for further studies.

## Methods

**An example of a training and assessment process.** An example role-playing session was developed to describe the training and assessment process. This role-playing session was used for preceptors in Saga University to conduct an assessment training session.

This is the first day that a resident works in an outpatient department. After seeing a patient, he presents a case to a trainer in charge of the day. The task for a trainer is to listen to the case presentation, analyze the diagnostic reasoning ability of the trainee, and give feedback to him.

Trainee: I will present a case, OK?

Trainer: Yes, please.

Trainee: Patient is 41-year-old female, whose chief complaint was general fatigue. No other symptoms were notified. I asked her if her interests are decreased but she denied. In the health check up two years ago, she has told nothing was wrong. Physical examination, routine laboratory, and physiological tests added no more information. As an assessment, she does not have any evidence of a disease.

Trainer: All right. What do you think are her differential diagnoses?

Trainee: Oh, well. I thought of depression is an important diagnostic hypothesis but she did not mention any decreased interest. Otherwise, I did not have any other differential diagnosis.

Trainer: I see. Why do you think the patient has depression or any other thoughts?

Trainee: Umm, I remember that depressive patients complain of general fatigue. I have not seen any patient with general fatigue before.

Trainer: OK. You pointed out that her chief complaint was general fatigue. Depression is one of the KDDs for a general fatigue patient. That's fine. However, still your process of information gathering is not systematic. You did not mention when the symptom started or in what way. Did it start suddenly or gradually?



Trainee: I did not ask that information.  
Trainer: It is good that you honestly told me what you did not do. Anyway, onset of a symptom is very important information because it sometimes relates to pathological causes. Next, what about other differential diagnoses?  
Trainee: How about anemia or hypothyroidism?  
Trainer: Very good! If you think the patient is anemic, what kind of question do you want to ask?  
Trainee: If a patient has tarry stool, GI bleeding might be the cause of the anemia.  
Trainer: Yes. When you think the patient has hypothyroidism, what kind of physical finding is important?  
Trainee: Oh, I did not check her goiter.  
Trainer: OK. So, let's see the patient together.

In this dialog, the trainer gave a score of 2 using the assessment scheme in Table 1. She thought that the trainee's experience with undiagnosed patients was quite limited. She tried to give him positive feedback to facilitate the retrieval of more differential diagnoses and to let him know how to connect each diagnostic hypothesis with history taking and physical examination (H&P).

### How to Use the Initial GRS in Actual Training and Assessment

**Settings of clinical training.** If planning to use the GRS model in clinical settings, the first thing to consider is the level of clinical training. Since it is expected that this model is used in the process of diagnostic reasoning, a clinical trainee sees an undiagnosed patient, sorts out case information while considering diagnostic hypotheses, and then presents the case to a trainer. Thus, it is necessary for the trainee to see patients mainly in an emergency or outpatient department, with or without supervision.

It is possible to use the model in a ward, but the trainee has to see a patient without knowing the diagnosis or any other information related to the patient. Even if such a condition is maintained, diagnostic reasoning training in a ward is still not as useful as in an emergency or outpatient department because some patients already know their diagnoses and have become accustomed to being asked their histories.

**Case selection.** When you use the GRS model, some cases should be excluded from the assessment. First of all, a prerequisite for this model is that a trainee sees an undiagnosed patient. From our experience, patients should be excluded if their cases do not follow common processes of diagnostic reasoning, such as those seeking medical certificates or prescriptions only, those referred from other facilities, and those who want secondary health checkups or second opinions.

Moreover, elderly patients tend to have multiple morbidities. When a patient has already been diagnosed with several

diseases and presents with a new symptom or problem, it is still possible to use the GRS model but the case difficulty must be considered high.

**Settings for assessment.** In outpatient training settings, H&P usually precedes assessment and decision for therapy or management. Novice trainees are usually told to present the case just after H&P without letting the patient go home, because trainers must pay attention to the accuracy of diagnoses and management for patients. In this situation, as in the role-playing example above, the trainer can use the GRS model and then see the patient with the trainee afterward. In this setting, assessment is carried out in a one-to-one teaching situation.

More experienced trainees might be able to see patients independently. In that case, trainers will facilitate a reflective assessment session for multiple cases after the trainee finishes seeing all the patients. Assessment in a one-to-one teaching setting is normally used, but a trainer can also accommodate multiple trainees or multiple trainers can assess one trainee.

### Pilot Study

From June to July 2009, 17 fifth-year medical students saw 84 patients in the General Medicine Outpatient Department of the Faculty of Medicine Hospital at Saga University. Ten preceptors in the General Medicine Department asked the students to see the patients independently, listened to case presentations, and assessed each student using the scale in Table 1. Before case presentation, students were allowed to review the medical interview session for 10–20 minutes as preparation time. The students gave their consent to participate in this research.

The criteria for patient selection were those who came for the first complaint in the Outpatient Department and who agreed to consultation by a medical student as part of outpatient training. Patients seeking a second opinion or secondary checkup after an initial health checkup were excluded. Patients who came to the Outpatient Department with a referral letter from other clinics or hospitals were also excluded.

After analyzing descriptive statistics, univariate analyses conducted were *t*-test for difference between student genders and analysis of variance for difference among different students and faculty. Alpha error was set at 5%.

### Construction of the New Four-point GRS

To construct a four-point GRS from the original five-point GRS, the following two processes were executed: defining the theoretical background of the GRS and relating the GRS with its convergent model of RIME. Since RIME is a more comprehensive assessment than diagnostic reasoning, each level of RIME is compared with the new four-point GRS from the viewpoint of diagnostic reasoning.

Unless a presentation is assessed at the highest level, feedback should be given with the aim of helping the trainee



advance to the next level. Recommended feedback is attached with each level of the assessment.

Crossley and Jolly<sup>10</sup> described four general principles required for WBA, and the newly developed GRS reflects these four principles.

## Results

**Pilot study.** For the cases of 83 patients, results of WBA by GRS are described in Table 2. Under assumption of interval scale, the mean score was 3.34 and standard deviation was 0.73. Each faculty member assessed 5–13 students. The average scores of faculty members significantly differed ( $P < 0.001$ , range: 2.87–4.25). Each student saw two to seven patients. The average scores of students ranged from 3.00 to 4.00 without statistical significance ( $P = 0.52$ ).

No student was given a score of 1. One reason for no score of 1 is that only a few students were directly supervised during the actual consultation process. Relationship between actual H&P of a case and its case presentation remains unclear.

This result led to two changes in the initial assessment model. The first change was to decrease the five-point GRS into a four-point GRS, because no trainer gave a score of 1 to any trainee. The other change was to add a descriptor for each level of the GRS, because some trainers felt that descriptors would make it easier to understand and memorize different levels of case presentations.

**Table 2.** Results of pilot study.

ASSESSMENT RESULT	# OF CASES	STUDENT AVERAGE	SCORE
2	4	S01	3.20
3	56	S02	3.20
4	14	S03	3.71
5	9	S04	3.17
		S05	3.75
		S06	3.00
		S07	3.20
		S08	3.00
		S09	3.50
		S10	4.00
		S11	3.67
		S12	3.00
		S13	3.33
		S14	3.50
		S15	3.00
		S16	3.00
		S17	3.20

  

DESCRIPTIVE STATISTICS	
Mean	3.34
Standard deviation	0.73

  

FACULTY AVERAGE	SCORE
A	3.80
B	4.25
C	3.90
D	2.87
E	3.00
F	3.00
G	3.00
H	3.33
I	3.15
J	3.20

## Confirmation of Logics for the New Four-point GRS

**Vague presentation (V-level).** Some case presentations by medical students are assessed as vague because of insufficient or unordered information in the presentation. There are two main reasons for vague presentations, limited training for case presentations and limited clinical experiences.

If a patient complains of pain in the beginning of a medical interview, even a student without any diagnostic hypothesis is able to ask questions by following the OPQRST scheme (onset, provocation or palliation, quality, region and radiation, severity, and time course), perform screening physical examinations, and sort out the collected information into a case presentation. This is the most basic mechanism of case presentations at the reporter level in the RIME model. If a student cannot complete the relevant clinical tasks, the information they gather will be vague, and their case presentation will also be vague.

Tham<sup>11</sup> proposed an ORIME model for assessment in clinical training in Singapore. The O added to the RIME model stands for *observer*. Vague presentations mean that presenters do not play the role of reporters and are, therefore, assessed as *observer* in the ORIME model and as V-level in this study. Since such observer trainees do not contribute to any clinical service provision, it would be difficult to call such training clinical clerkship.

**Structured presentation (S-level).** In the RIME or ORIME model, whether trainees can play the role of reporters is crucial. If trainees can take complete H&P information from the patient and deliver it to senior physicians, those trainees can help senior physicians check clinical information. Such trainees are expected to work as clerks, gathering information systematically from patients and delivering it to the trainers. Such trainees do not attain the level of interpreting H&P information with regard to differential diagnoses.

From a trainer's viewpoint, structured presentation means that the information taken from the patient is routinely complete and that the structure of the case presentation is satisfactory. When a trainer listens to a structured presentation, he/she should find logical flow, accuracy of H&P information, and comprehensiveness of routine H&P.

If a trainer finds any problem in logical flow and has difficulty in understanding the case, more practice is needed in H&P and case presentation. If a trainer feels that his/her own H&P information is different from the trainee's at a basic level, he/she should check if the information is accurate. S-level trainees do not have to list KDDs, but should cover routine H&P and deliver the information to the trainer.

**Organized presentation (O-level).** Trainees in this level are able to list most KDDs but still have difficulty collecting pertinent positive or negative signs and symptoms. The trainee is able to utilize explicit H&P information well enough to list KDDs. O-level trainees are able to conceive most KDDs from the H&P information but still have difficulty asking questions or selecting focused physical examinations in order to collect pertinent positive and negative signs and symptoms.





Croskerry's dual processing theory<sup>12</sup> advises diagnosticians to utilize both intuitive type 1 and analytic type 2 cognitive processes. Expert diagnosticians are able to list most KDDs using a type 1 process and draw on the framework of those KDDs to ask questions in a type 2 process. O-level trainees show insufficiency in either type 1 or type 2 processes compared to an expert diagnostician.

Still, O-level physicians are able to list most KDDs from the information collected in a medical interview or physical examinations. In that sense, differential diagnoses and H&P information are organized into the case presentation. Since coverage of pertinent signs and symptoms relevant to each KDD is not complete in this level, trainers should ask O-level trainees to use each KDD to think of any pertinent signs and symptoms. To help O-level trainees learn to ask questions in order to collect pertinent positive/negative signs and symptoms, trainers should ask about any unconsidered signs or symptoms that might rule in or rule out competing KDDs.

**Pertinent presentation (P-level).** To treat patients effectively, sound judgment must be used to guide decisions for management. Trainees should have KDDs in mind and use them help deductively retrieve signs and symptoms to include or exclude competing diagnostic hypotheses. If a trainer feels that a case presentation contains the relevant KDDs and all pertinent positive/negative information, such a presentation is assessed as pertinent.

Here, the previous five-point GRS has been changed into a new four-point GRS with levels of vague, structured, organized, pertinent (VSOP, Table 3). Hereafter, the four-point GRS is referred to as the VSOP model.

## Recommended Feedback to Trainees in Each Level

**Vague presentation (V-level).** When a preceptor listens to a vague case presentation, one or both patterns of circumstances are often at play: incompleteness of a presenter's H&P

or poor skills in case presentation. Differentiating these two patterns is important to decide which feedback a preceptor should give to the presenter.

Some presenters had obtained complete H&P information but failed to include such information in the presentation. To differentiate categories of vague presentation, the first question to ask is "Do you know which information should be included in the presentation?" If the presenter's understanding is OK, the trainer will discuss the incorrect part of the case presentation. When presenters' H&P skills are identified as incomplete, trainers should specify which skills are incomplete and give specific advice.

The main difference between vague and structured presentations is completeness of information gathering for screening. Since trainees in this level cannot systematically list differential diagnoses, completeness refers not to pertinent positives and negatives but to routine information.

**Structured presentation (S-level).** When the presentation contains all the screening H&P information, it means the presentation is structured. The main difference between structured and organized presentations is whether KDDs are included or not. It is supposed that S-level presenters are too occupied with complete screening H&P to think of KDDs. Trainers should ask the presenters to think of KDDs to guide S-level presenters up to O-level.

Frequently, S-level trainees have their own diagnostic hypotheses but cannot list many KDDs. Positive feedback should include completeness of screening H&P information and several KDDs. As corrective feedback, trainers should ask if the trainees can think of any other KDDs. In case trainees do not think of any KDD, trainers should ask questions to narrow down anatomical systems or pathophysiological causes of specific differential diagnoses.

If the case presentation is about a common type of chest pain and differential diagnoses are only killer four chest pain

**Table 3.** Revised global rating scale for diagnostic reasoning.

CONDITION OF PRESENTATION		LEVELS OF THE PRESENTER	FEEDBACK FROM THE TRAINER TO THE PRESENTER
Vague	Vague presentation due to insufficient or unordered information or poor expression of the contents.	Gathering or organization of information is not systematic. Practice for case presentation is needed.	Point out what is missing in case presentation. Tell the trainee to have one-to-one practice.
Structured	Case presentation is structured with routine H&P information is but key KDDs are still lacking.	Able to report the case information but unable to list all KDDs by interpreting the patient's problems.	Give positive feedback for complete information. Ask the presenter to think of any more KDDs. If not, the presenter should be recommended to narrow down the extent of differential diagnoses.
Organized	Most KDDs are listed and organized with H&P information in case presentation, but pertinent positive and negative S/S are insufficient.	During H&P no relevant S/S to KDDs are obtained.	Give positive feedback for KDDs and corrective feedback for pertinent positive and negative S/S.
Pertinent	Pertinent positive and negative S/S relevant to KDDs are covered.	Through H&P whole picture of the case and its KDDs are clearly described.	Give positive feedback for a good presentation. Ask the presenter to specify the lessons learned from the case.

**Abbreviations:** H&P, history taking and physical examination; S/S, signs and symptoms; KDD, key differential diagnoses.



diseases (acute coronary syndrome, aortic dissection, pulmonary embolism, and tension pneumothorax), the trainer should ask if the presenter can think of any more common diseases. Some S-level presenters cannot balance both severe diseases to be excluded and more commonly occurring diseases.

**Organized presentation (O-level).** Case presentations by O-level trainees contain complete H&P and many KDDs, but lack a few KDDs and pertinent positive and negative signs and symptoms. As positive feedback, trainers should tell trainees that H&P are complete and many KDDs are organized into the case presentation.

As corrective feedback, trainers should first ask if O-level trainees are able to think of any more KDDs. If not, trainers should advise trainees to narrow down differential diagnoses. Finally, trainers might check if trainees know questions for pertinent positive and negative symptoms or physical examinations for pertinent positive and negative signs. If not, trainers should tell the trainee how to check for such pertinent information.

**Pertinent presentation (P-level).** Normally, only positive feedback is given for P-level trainees, because no more improvement is needed from a diagnostic viewpoint. It would also be recommended to ask the presenter to specify the lessons learned from the case. Further discussion might be possible for therapeutic or management purposes.

#### Four General Principles of Work-based Assessment

Crossley and Jolly<sup>10</sup> pointed out that WBA shows poor engagement and disappointing reliability and suggested four general principles to overcome those issues. Here, we consider the VSOP model with regard to their four general principles. First, the response scale should be aligned with the reality map of the judges. In VSOP model, trainers have to listen to a trainee's case presentation, grasp the whole picture of the case, and assess it. This is the same as a daily precepting process and is therefore aligned to the judges' reality.

Second, judgments rather than objective observations should be sought. This WBA tool is not used for in vitro settings such as an objective structured clinical examination (OSCE), but for the in vivo practice setting.

Third, the assessment should focus on competencies that are central to the activity observed. Many preceptors feel that case presentation is useful for assessment in outpatient settings. This WBA will help trainers more objectively and actively assess trainees. The focus of this WBA is exactly on the competency of diagnostic reasoning.

Finally, the assessors who are best placed to judge performance should be asked to participate. For this WBA those who have already been teaching will be the assessors, as they have the best knowledge of the trainees, including their performance and progress. Based on these analyses, this WBA tool satisfies the general principles of a good WBA for use in clinical settings.

#### Discussion

The initial model of GRS for diagnostic reasoning showed variation in assessment results for trainers and trainees. Trainers had statistically significant variation caused by incomplete understanding or definition of GRS. Trainees did not show significant variation because they had seen only a very limited number of cases or had no training in settings other than the medical school.

The VSOP model, revised from the initial GRS, is the first WBA model for diagnostic reasoning using case presentations. The confirmation processes are logical and easy to understand from clinical viewpoint. Validity was checked using Crossley's four general principles of WBA.

#### Directions for Further Study

**Combination with relevant teaching models.** The VSOP model has a rating scale to measure achievement of diagnostic reasoning in clinical practice, but does not provide ways to ask questions to probe a trainee's understanding of diagnostic processes. Both the One-minute Preceptor and SNAPPS are popular models used in precepting in outpatient education, as they combine case presentation and one-to-one teaching with guidance on how to ask questions. If a trainer does not know how to probe a trainee's understanding, the One-minute Preceptor or SNAPPS can be combined with the VSOP model.

**Practice settings and specialty.** The VSOP model was developed as a tool that can be used in any setting and any specialty. Facilities may vary from a large hospital to a small clinic. Specialties also may vary from general practice to highly specialized areas. If we consider population-based health care, morbidity prevalence may vary depending on the settings. Trainers in each institution or each specialty must have different mental models of KDDs to assess the diagnostic reasoning ability of trainees in keeping with the demographics of different settings. It is expected that confirmation studies would be conducted in various settings.

**Case difficulty.** As a diagnosis of each case is the main focus of the VSOP model, we should notice that case difficulty may vary. It is natural to think that uncommon diseases are more difficult to diagnose than common diseases. A new problem emerges in multimorbidity, and these cases should also be regarded as difficult.

Issues in communication with patients can also increase the difficulty of a case. Some patients may be cautious about establishing relationships with medical students or residents, incoherent in their dialog with the trainee, or burdened by issues of cognitive ability. Some patients, especially in pediatrics or gerontology, are represented by family members, who will be the main target of the medical interview. Case difficulty in such cases will be higher.

**Validity study.** Alves de Lima et al<sup>13</sup> evaluated validity for Mini-CEX by examining if the instrument was capable of discriminating between pre-existing levels of clinical

seniority. Since both Mini-CEX and the VSOP model use GRSs, a similar study should be conducted to seek evidence regarding validity.

**Reliability study.** Reliability of an assessment tool is only one part of validity but it becomes very important when the tool is used for a pass–fail decision.<sup>14</sup> Alves de Lima et al<sup>13</sup> also checked the reliability of Mini-CEX using generalizability theory. The same kind of analysis can be applied to confirm the reliability of the VSOP model.

Alves de Lima et al<sup>13</sup> used a nested design of generalizability theory; since all the cases that trainees see are different, the universe of the trainer and the universe of the trainee are nested in the universe of the case in settings where one-to-one teaching is the main strategy. Furthermore, when several trainers listen to a trainee's case presentation, such as the setting of a case conference, analysis by crossed design of cases and trainers (assessors) is also possible.

If OSCE is conducted, multiple trainees are assessed on the same case, because standardized patients can simulate a case repeatedly. In this format, analysis by crossed design of cases and trainees is possible. OSCE is not a WBA, but could be used to conduct a study for reliability.

## Conclusion

The initial model of the five-point GRS for diagnostic reasoning showed variation in assessment results for trainers and trainees and was changed into the four-point VSOP model. This model is the first WBA for diagnostic reasoning based on trainers listening to case presentations. Validity was checked using Crossley's four general principles of WBA, but there are many issues for further study.

## Acknowledgments

I would like to express appreciation for Dr. Yasutomo Oda in Saga University because he provided the opportunity of this pilot study.

## Author Contributions

Literature search, tool development, data analysis and manuscript writing: HO. The author reviewed and approved of the final manuscript.

## REFERENCES

1. Onishi H. The role of case presentation for teaching and learning activities. *Kaohsiung J Med Sci*. 2008;24:356–360.
2. Neher JO, Gordon KC, Meyer B, Stephens N. A five-step “microskills” model of clinical teaching. *J Am Board Fam Pract*. 1992;5:419–424.
3. Wolpaw TM, Wolpaw DR, Papp KK. SNAPPs: a learner-centered model for outpatient education. *Acad Med*. 2003;78:893–898.
4. Elstein AS, Shulman LS, Sprafka SA. *Medical Problem Solving: An Analysis of Clinical Reasoning*. Boston: Harvard University Press; 1978.
5. Gruppen LD, Palchik NS, Wolf FM, Laing TJ, Oh MS, Davis WK. Medical student use of history and physical information in diagnostic reasoning. *Arthr Rheum*. 1993;6:64–70.
6. Pangaro L. A new vocabulary and other innovations for improving descriptive in-training evaluations. *Acad Med*. 1999;74:1203–1207.
7. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65:S63–S67.
8. Norcini JJ, Blank LL, Duffy D, Fortna GS. The mini-CEX: a method for assessing clinical skills. *Ann Intern Med*. 2003;138:476–481.
9. Onishi H. Assessment methods for clinical reasoning. *Nihon naika gakkai zasshi (Jpn J Intern Med)*. 2008;97:2596–2603.
10. Crossley J, Jolly B. Making sense of work-based assessment: ask the right questions, in the right way, about the right things, of the right people. *Medical Education*. 2012;46:28–37.
11. Tham KY. Observer-Reporter-Interpreter-Manager-Educator (ORIME) framework to guide formative assessment of medical students. *Ann Acad Med Singapore*. 2013;42:603–607.
12. Crosskerry P. Clinical cognition and diagnostic error: applications of a dual process model of reasoning. *Advances in Health Sciences Education*. 2009;14:27–35.
13. Alves de Lima A, Barrero C, Baratta S, et al. Validity, reliability, feasibility and satisfaction of the Mini-Clinical Evaluation Exercise (Mini-CEX) for cardiology residency training. *Med Teach*. 2007;29:785–790.
14. Downing SM. Reliability: on the reproducibility of assessment data. *Med Educ*. 2004;38:1006–1012.