Challenges and solutions in developing an objective and structured clinical examination for complementary and integrative medicine: a qualitative study

Aram Jeong  
Gachon University

Hye-Yoon Lee  
Pusan National University

Sangwoo Shin  
Pusan National University

Sunju Im (✉️ sunjuim11@hanmail.net)  
Pusan National University School of Medicine

Research Article

Keywords: Objective structured clinical examination, Qualitative analysis, Module development

Posted Date: July 7th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3090056/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

Background

The growing prevalence of chronic diseases emphasizes the importance of multidisciplinary and integrative medical care, which considers various factors in the diagnosis and treatment processes. Therefore, training and evaluation on information gathering, physical examination, and patient education for ideal integrative medical care are necessary. An objective and structured clinical examination (OSCE) is widely used in medical education as a tool for evaluating overall clinical performance. This study developed OSCE modules for ideal complementary and integrative medical care practice in Korea. We report the problems and solutions that occurred during this process, as well as future tasks.

Methods

A total of 21 OSCE modules were developed according to 3 different diseases for each of the 7 clinical presentations (CP). Seven clinical experts developed the OSCE modules in each specialized field. Quality control was conducted through repeated feedback from two medical education experts and a standardized patient educator (SP educator). Analysis of the contents of each feedback, a survey of the 7 clinical experts, weekly meetings, and a focus group interview (FGI) was conducted to derive the challenges and possible solutions. Self-evaluation of OSCE development competency and importance-performance analysis (IPA) were conducted for the 7 experts after the main development process.

Results

Seven main themes and 18 subcategories were extracted. The main challenges of developers were categorized into “case,” “test situation,” “post-encounter note,” “checklist,” “scenario,” “format,” and “pattern identification.” During module development, they solved these challenges through discussions among developers and medical education experts. All solutions were categorized into 33 codes. Our survey found their competency in all items to be better than before development, and that they considered pattern identification (PI) the greatest challenge due to its ambiguity.

Conclusion

We found that the more OSCE modules the developers worked on, the more their competency was improved. However, they faced many challenges when developing the modules, which they resolved through discussions on the project. For further effective OSCE module development, we note that social and college-level support should be provided in the form of standardized schemas and human and spatial resources.
Background

Awareness on the importance of providing integrative medicine has increased worldwide. With the growing prevalence of lifestyle-related chronic diseases and degenerative diseases brought about by aging, a multidisciplinary approach to treatment is necessary [2, 3]. In the field of treatment, perceptions of the patient-doctor relationship are constantly changing, and these changes must be reflected in training and evaluation [1].

The process of Korean medicine (KM) practice involves using clinical reasoning based on patient information to further inform diagnostic processes and provide guidance on lifestyle management and treatment plans [4]. KM treatments are more individualized compared to those of conventional medicine, which is known to rely on “pattern identification” (PI) [5]. PI involves subdividing present symptoms to classify aspects of the same diagnosis, subdividing the transition state of diseases or comprehensively assessing personal factors, such as patient’s usual physical condition and lifestyle, and nature and location of illness. Even though two patients have the same disease, their PI can be different [6]. Therefore, at the diagnosis stage, it is needed for KM physicians to improve their competency in PI diagnosis interference.

The objective and structured clinical examination (OSCE) is a tool for assessing students’ clinical performance competencies, including history taking, physical examination, patient education, and patient-physician interaction (PPI) when they treat a standardized patient (SP) in simulated medical conditions [3, 7, 8]. As OSCE values authenticity consistent with the medical field, its main goal is to train and assess the essential competencies in an environment similar to the actual field [9].

Regarding integrative medicine, conventional medicine and KM operate as dual healthcare systems. While the two overlap by making diagnoses using the same disease classification code, the treatment methods used in KM are unique, such as herbal medicine, acupuncture, moxibustion, and manipulative therapies like Chuna. Since much of KM care is performed as a form of primary care, an OSCE is an adequate training and assessment tool for KM care competencies as it provides an initial diagnosis, diagnosis and treatment plan, and essential patient education in the current state [10].

It is necessary to develop OSCE modules that reflect KM characteristics, especially PI reasoning with diagnostic reasoning, planning, education, and explaining in terms of the KM principle [11]. A closer examination and critical analysis of the types and quantity of information to obtain, and procedures for history taking and physical examination, are essential in ensuring the required information is secured. OSCE modules that can fully reflect these KM characteristics and are suitable for student training and assessment are therefore necessary. This study analyzed and reports the problems and solutions that occurred in this process.

Methods

2.1. Setting
This study looked into the difficulties and problems experienced by OSCE module developers during the development process. First, we analyzed the minutes of meetings during the development process. Second, a survey of OSCE module developers was conducted at each meeting. Third, we conducted a focus group interview at the last meeting, after all module development was completed.

This study was conducted through online Zoom meetings and Google Forms surveys because of the COVID-19 pandemic. The National Institute for Korean Medicine Development funded the OSCE module development as part of the “Development of OSCE for internal medicine disease.” In this project, the developers were required to develop OSCE modules for the following seven internal diseases: “hypertension,” “functional dyspepsia,” “cancer-related symptom relief,” “Parkinson’s,” “migraine,” “vertigo,” and “chronic fatigue.” The module items that the developers were required to create were “schemas,” “scoring criteria,” “case management tables,” “door instructions,” “post-encounter notes,” “checklists,” and “scenarios for SP training.” OSCE development was performed by the developers from Sep 1 to Oct 31, 2021, and during this period, they created three modules for each disease. Whenever the developers completed each module, it was reviewed by a medical education expert with 17 years of experience (the corresponding author of this study, Im S) and an SP trainer with 8 years of experience. The reviewed comments were provided to the developers in the form of written feedback.

### 2.2. Participants

The participants were all faculty members who had participated in the project “Development of an educational tool (OSCE) based on the KM standard clinical practice guideline (CPG).”

In this study, a total of 10 faculty members were included: 7 researchers participated in the development of the OSCE module and 3 experts reviewed the created OSCE modules (1 KM education expert, 1 medical education expert, and 1 SP trainer). The OSCE developers were all clinical faculty members from the College of KM who volunteered to participate in this project.

The seven developers who specialized in oriental internal medicine ($n = 4, 57\%$), Sasang constitutional medicine ($n = 2, 29\%$), and oriental neuropsychiatry ($n = 1, 14\%$) included 4 men and 3 women. They have an average of 6 years of clinical experience and 10.8 years of medical education experience (Table 1).

All participants took part in 1) the survey, 2) the meetings, and 3) the focus group interview, while two of them were included in the rest of the analysis except for the surveys. All of the participants agreed to participate voluntarily and have the meetings recorded and transcribed for this study.
Table 1
Participants’ baseline characteristics

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Gender</th>
<th>Specialization</th>
<th>Education career (years)</th>
<th>Years of clinical experience</th>
<th>Experience in developing OSCE</th>
<th>Experience in implementing OSCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>40s</td>
<td>F</td>
<td>IM</td>
<td>6</td>
<td>3</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>B</td>
<td>40s</td>
<td>F</td>
<td>IM</td>
<td>11</td>
<td>16</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>C</td>
<td>30s</td>
<td>M</td>
<td>IM</td>
<td>4</td>
<td>14</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>D</td>
<td>30s</td>
<td>F</td>
<td>SC</td>
<td>6</td>
<td>10</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>E</td>
<td>40s</td>
<td>M</td>
<td>SC</td>
<td>5</td>
<td>8</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>F</td>
<td>30s</td>
<td>M</td>
<td>IM</td>
<td>2</td>
<td>10</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>G</td>
<td>40s</td>
<td>M</td>
<td>NP</td>
<td>8</td>
<td>15</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

F, female; M, male; IM, internal medicine; SCM, Sasang constitutional medicine; NP, Neuropsychiatry; OSCE, Objective Structured Clinical Examination

2.3. Data collection and analysis

2.3.1. Survey

A survey was conducted at every meeting held for this project. The questionnaire consisted of the following items: (1) the basic characteristics of the developers (age, gender, specialization, education experience, clinical experience, experience developing OSCE modules, and experience of delivering OSCE education to students; Table 1), (2) The change of their self-competency during developing OSCE modules: “In developing OSCE modules, I can select adequate cases,” “I can write door instructions,” “I can write the interim test,” “I can create the checklist,” “I can make scenarios for standardized patients,” and “I can develop cases for the OSCE.” The statements were rated by the participants on a scale of 1 to 5, ranging from very poor to very good. We conducted the questionnaires at the beginning of every meeting (Fig. 1).

2.3.2. Analysis of the meeting minutes

A total of five meetings were held (Sep 01, 2021; Sep 15, 2021; Sep 29, 2021; Oct 13, 2021; Oct 27, 2021). The meetings took between 75 and 90 mins and were recorded online using the Zoom platform by a single researcher (HY Lee), who then transcribed the audio recordings. The participants were allowed to speak spontaneously about their difficulties and problems.

At the beginning of the meeting, the SP trainer and the first authors (HY Lee and AR Jeong) conducted a demonstration based on the scenarios produced by the developers under the supervision of the main author (S Im). Following the demonstration, the researchers reviewed each scenario from the perspective of the scorer, identifying any problems and discussing solutions.
After the meeting, the authors individually coded the data using a qualitative descriptive approach. We tried to describe the participants’ experiences in easily understood language. The coding was carried out by three authors (HY Lee, AR Jeong, and S Im) and checked whenever the author’s meetings were finished. A total of six additional meetings between the authors were additionally held (September 24; October 1, 8, 15, and 22; and November 5, 2021) to review and verify the codes.

To increase the reliability, we mainly collected codes completed by each author with more than 85% agreement between coders. We used a conventional coding method that flowed directly from the data and an in vivo coding method where the expressions of the participants were directly quoted as much as possible. After conducting all coding, repeated and dominant codes were categorized into main themes along with the participants’ expressions. During this process, we double-checked that the codes corresponded to the participants’ intentions. Then, we collected the codes by using lean coding to reduce the number of codes by grouping similar codes into themes. Finally, categories and subcategories were defined around the difficulties, challenges, and solutions of OSCE module developers (Table 2).
<table>
<thead>
<tr>
<th>Theme</th>
<th>Challenges</th>
<th>Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Case</td>
<td>Difficulty in case selection</td>
<td>Cases should be</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- common, typical, or clinically important.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- diverse in age and gender.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- related to the CPG.</td>
</tr>
<tr>
<td>2 Test</td>
<td>Authentic case (i.e., using cards for physical examination)</td>
<td>Cases should be authentic.</td>
</tr>
<tr>
<td>situation</td>
<td>Lack of critical patient information</td>
<td>Essential information should be provided in advance.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>e.g., diagnosed with cancer, medical history sheet for chronic diseases, prescriptions, body mass index.</td>
</tr>
<tr>
<td>3 Post-</td>
<td>The difficulty of setting the scoring criteria and format</td>
<td>Scoring criteria need to be agreed upon in the actual test situation by the interrater.</td>
</tr>
<tr>
<td>encounter</td>
<td></td>
<td>The item number of patients’ education and future plans should be adjusted.</td>
</tr>
<tr>
<td>note</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Checklist</td>
<td>Too many items</td>
<td>Items should focus on</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- a high-level differential diagnosis according to the schema.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- essential contents that are related to the cases (i.e., case-specific past and social history).</td>
</tr>
<tr>
<td></td>
<td>Missing critical differential diagnosis</td>
<td>Essential items including history taking and PE for differential diagnosis should not be omitted.</td>
</tr>
<tr>
<td></td>
<td>Difficult to score</td>
<td>Items in the checklist should</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- be observable.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- set a clear scoring standard.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- avoid using terminology or Chinese characters to help SPs and scorers understand.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- be simple to improve readability.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- avoid duplication.</td>
</tr>
</tbody>
</table>

SP, standardized patient; CPG, clinical practice guideline; PE, physical examination; OLD CoEx CAFÉ; onset, location, duration, course, experience, character, associated symptom and factor; PI, pattern identification
<table>
<thead>
<tr>
<th>Theme</th>
<th>Challenges</th>
<th>Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The order of items is different from that of execution</td>
<td>The order of items should be in the order - generally done in clinical situations - from head to foot.</td>
</tr>
<tr>
<td>5  Scenario</td>
<td>Insufficient information</td>
<td>Sufficient personal information should be provided including emotional status and current situation.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sufficient information about general health condition should be provided including social history and broad information for syndrome differentiation.</td>
</tr>
<tr>
<td></td>
<td>Difficult for SP to understand</td>
<td>Terms should be easy and should avoid using terminology or Chinese characters.</td>
</tr>
<tr>
<td></td>
<td>Concerns about differences between SPs within the same case</td>
<td>Specify when to ask the question if there are “must-ask questions“ Delete questions that cannot be standardized Provide clear directions for SPs’ action</td>
</tr>
<tr>
<td></td>
<td>SPs’ dialogue includes critical contents</td>
<td>Critical contents can only be provided when the exact question is asked by students.</td>
</tr>
<tr>
<td></td>
<td>Limited range of patients’ education</td>
<td>Patients’ education can largely concern future diagnosis procedures when the diagnosis is not confirmed</td>
</tr>
<tr>
<td>6  Format</td>
<td>Out of format</td>
<td>Developers need to understand and be familiar with the format. Items should fit the section, simplify the present illness. The length should be possible to memorize.</td>
</tr>
<tr>
<td></td>
<td>Inconsistency in the contents</td>
<td>The contents of the checklist, scenario, and present illness should be consistent.</td>
</tr>
<tr>
<td>7  Pattern identification</td>
<td>The gap between schemas and real clinical situations</td>
<td>Researchers’ agreement is required in using schema reasoning training for student education and evaluation.</td>
</tr>
<tr>
<td></td>
<td>Uncertainty of the existing PI schema</td>
<td>The PI schema should be developed i.e., categorizing each related PI, checking consensus between researchers</td>
</tr>
</tbody>
</table>

SP, standardized patient; CPG, clinical practice guideline; PE, physical examination; OLD CoEx CAFÉ; onset, location, duration, course, experience, character, associated symptom and factor; PI, pattern identification
### 2.3.3. Questionnaire for an importance and performance analysis

At the final meeting (Oct 27, 2021), we conducted another survey to measure awareness of the importance of one's ability to perform and self-evaluation using a questionnaire comprising 18 items derived from the analysis of the meeting minutes. The awareness of the importance of one's ability to perform and self-evaluation were scored on five-point Likert scales (5 = very important for importance and 5 = excellent respectively for performance). This survey was intended to identify what the developers needed for further effective OSCE module development. In addition, they checked the importance and performance of each item in the questionnaire. The importance-performance analysis (IPA) analysis method was used to examine the final survey results, which were then plotted using the Excel application (Fig. 2). IPA analysis has been mainly used in service marketing to measure customer satisfaction; however, it has been also applied to different areas such as the competencies of college students, faculty members, and researchers [12, 13]. All OSCE module developers took part in the surveys.

### 2.3.4. Focus group interviews

To acquire more data, at the last meeting (Oct 27, 2021), all participants took part in an interview conducted by the first author (AR Jeong) and recorded with a voice recorder; it was then transcribed immediately afterwards by the other authors. For greater validity, prior to the interview, our final coding was presented to all participants to solicit their opinions and feedback (member checking). The interview lasted about 40 minutes and took the form of a focus group. The interview questions were as follows: “What are the good points in developing OSCE modules?” “What difficulties have you had in developing OSCE modules even though you received feedback from specialists?” and “What kind of help do you need from the college and society to develop OSCE modules and provide OSCE education?” Following the interview, the authors double-checked to see if the interview results differed from our initial coding.

### 2.3.5. Member checking

To maintain validity, we recorded what the participants said at the meetings and interviews during the study period and confirmed with each participant whether our recording was accurate.
Results

3.1. Baseline demographics of the participants

All seven developers (100%) participated in all surveys and meetings. They were all faculty members who had clinical experience of 3–16 years (mean: 10.8 years) and education experience of 2–11 years (mean: 6 years; Table 1).

3.2. Self-assessment of confidence in developing OSCE cases

The developers responded that during the OSCE case development, their competencies for all items of “case selection,” “situation guide,” “post-encounter note,” “score,” “scenario,” and “case creation” were better than before development (Figure 1). The average in the 1st survey was 18,857 points, the 2nd was 21.429, the 3rd was 23, and the 4th was 23.429 based on a total of four self-competence evaluations. As a result, we confirmed that the developers’ self-competence grew as the development process was repeated.

3.3. Challenges and solutions

Through an analysis of the meeting minutes, we identified the difficulties and problems the developers experienced and subcategorized them into challenges (Table 2). Based on the results of the analysis, 18 subcategories were extracted under 7 main themes.

3.3.1. Case

Most of the developers expressed that deciding the criteria to use to select a sample case was difficult. The main contents were as follows.

One participant said the following at the 2nd meeting:

“I am not sure what kind of criteria I should use when I choose a patient case. Given this is an educational module, I’m curious whether an extreme case is preferable over a common patient case. I am not sure which is the more important between OSCE development itself and the procedure of development.” (D1)

As a result of the meetings and feedback from medical education experts, the methods of solving this kind of challenge were classified into three subcategories, which were that the cases should be “common, typical, or clinically important,” “diverse in age and gender,” and “related to CPG.”

The medical education specialist’s feedback (S1):

“By simply changing the age and gender of the patient in each module, students can feel like it is a new case. And, because the purpose of OSCE is for evaluation rather than education, it is preferable to choose a case that is a case of a patient we see frequently in clinical practice.” (S1)
3.3.2. Test situation

The developers thought that it was very unnatural to present the results of oriental medicine examinations on a card in a test situation. For example, after the student performed a tongue and pulse examination on an SP, the result (for example, the tongue has coated a lot, the pulse is weak) was presented by the SP on a card.

“I'm concerned about how many times I should use the diagnosis card for writing tongue or pulse diagnosis results. I don't think it's a good idea to use it too much because it seems to be apart from reality.” (S1)

“I have thought about the problem a lot for a long time. As a result, I believe that it is best to use a model. For example, in the case of tongue diagnosis, it can be evaluated by making an artificial model with a photo attached to a face. However, this is difficult to apply because there is a problem with how realistically the photo resolution can be printed.” (S2)

Furthermore, according to the developers’ experiences, some necessary information is frequently missing in test situations. They believed that this problem should be solved. The following is an example:

“It appears that the SP must provide sufficient information about the case to the students in the test situation. For example, in the case of a 55-year-old woman diagnosed with cancer, information about medical history, previous diagnoses with cancer, or kinds of prescription should be presented to the student in advance…” (S1)

3.3.3. Post-encounter note

The developers suffered from a problem in presenting clear scoring criteria for the post-encounter note. Following a review by the medical education experts and SP trainer, the OSCE module was completed by the developers. It was determined that a common standard was required because each module was written based on a different standard. In addition, the expert and trainer suggested that the post-encounter note question and the score for each question should be adjusted based on the actual test situation. The following are some sample comments:

“In a OSCE situation, we should discuss the scoring criteria more, and in this project, there was no common scoring criteria. For example, some developers set the scoring standard as ‘2 points for including the first correct answer and 1 point for including the second correct answer,’ while others did not.”

They seemed to have difficulty suggesting scoring criteria in the post-encounter note. After the medical education experts and SP trainer reviewed their development, all researchers concluded that common score criteria for the post-encounter note are needed.
“It is better to show the number of right answers in the post-encounter note. And it should be numbered in order of the priority of the questions. For example, in the case of this kind of question, ‘Write down the examination plan,’ it is much better for the scoring criteria needed to be written as follows: 1) First priority right answer, 2) second priority right answer, and 3) third priority right answer.”

We should settle the scoring criteria when we perform a real OSCE. In this project, at the beginning, we did not standardize the scoring criteria. For example, it should be graded as 2 if the answer included the first priority answer, 3 for the second priority answer, and so on. Also, some differences were identified in the priority among the modules.

3.3.4. Checklists

The evaluation of OSCE is scored by the instructor and the SP based on the checklist. The checklist includes items such as taking the patient’s history, physical examination, patient education, and the patient-physician relationship (PPI). The following difficulties were identified by the developers when they created the OSCE checklists including score criteria. The specific contents were classified into four categories, as shown in Table 1.

3.3.4.1. Too many items

When the students performed the OSCE, the developers awarded points after checking whether the students asked the SP adequate questions. However, there were some problems regarding the number of checklist items, as there were too many questions in the checklist. This was caused by developing the OSCE based on the KM schema. The medical education experts proposed two subcategories for problem solutions. A representative difficulty and its solution are as follows:

“This is the problem I struggled with the most while creating the OSCE module. I had to derive both Conventional and traditional Korean medical diagnoses, so I had to add additional questions related to diagnosis, so the content was lengthy. So, I am not sure if I should omit some of the questions related to dialectics.” (D2)

“In KM clinical practice, the main difference between KM and Conventional medicine is performing its own diagnosis, so it is critical to use questions related to KM diagnosis effectively. So, it is crucial to include all kinds of items related to KM diagnosis when you develop the checklist. I hope that the sections dealing with Conventional medicine diagnosis will be reduced.” (S1)

“It is appropriate to have 10 to 15 items, focusing on the core history taking, and it's usually around 12 items. I'd like to put in a few more questions about KM diagnosis.”

3.3.4.2. Missing critical items

At the meetings, the researchers confirmed that the history taking or examinations required for differential diagnosis had been omitted from the checklist:
“When the patient says ‘My stomach hurts,’ according to the schema, it is necessary to distinguish whether it was induced by an ulcer, stress, drug, or heart disease, but the questions to distinguish it were missing.” (S1)

3.3.4.3. Ambiguous

After the developers observed the demonstration at the meeting, they discovered various problems with the checklist items. These kinds of challenge were categorized as “ambiguous.” First, there was difficulty for the scorer in evaluating whether the student was performing an adequate examination, such as inspection, and that is why scoring in a real OSCE situation may be impossible.

“It is difficult to check whether the students directly observe the patient’s lips, so it is preferable to remove the observation of lip color from the checklist.” (S1)

Second, there were numerous sections where the scoring criteria were presented in an ambiguous manner. For example, the item “Perfect physical examination” would be checked with one point, but the standards for performing it were not stated.

Third, the checklist had a problem in that, despite being scored by an SP with no medical education, the scoring criteria were written in difficult-to-understand terms using medical terminology or Chinese characters. Also, the scorers found that is it hard to understand due to the mix of oriental and Conventional medicine terms. The following is an example presented by an SP trainer who participated in the demonstration from the SP’s point of view:

“If you want to check somethings like ‘Did the student palpate the stomach area?’ please specify the location so that the SP can understand it easily.” (S3)

As a solution, it was suggested that the scoring criteria should only include items that can be observed in the actual test situation, and as far as possible, scoring criteria should be written in Korean so that scorers can easily understand them.

“If the drug name is Afatinib, please write it in Korean.” (S3)

Furthermore, some duplicated items were discovered in the checklist. In particular, developers were asked to classify the items related to patient education as presumptive diagnosis, presumptive dialectics, diagnosis plan, treatment plan, and education plan and not duplicate them.

“In case of dizziness, if the type of dizziness is important for evaluation, it should be evaluated separately as a different item. For example, how the dizziness changes with posture, whether the patient is conscious or not, and so on. When I reviewed the developers’ module, it was difficult to score because all of the items were combined into one item.” (S1)

3.3.4.4. The order of items is different from that of performance
The developers reported that it was difficult to check all of the contents quickly because the order of actual students’ performance order and the order of the checklist differed. As a result, there was some discussion about how the checklist should be ordered:

“The grading was difficult because the students’ questions did not appear in the order specified on the checklist. I had to score while searching for each question, and some of them were unclear, so how could I judge them?”

“So, when structuring the order of the questions, you prefer to create a question-centric framework so that students will be likely to ask the important questions more frequently. Also, it would be better to write the physical examination in the order of head-to-toe.” (S1)

3.3.5. Scenario for SP training

3.3.5.1. Insufficient information

The medical education experts and SP trainer reviewed the scenario and discovered that information about the patient’s emotional state, situation, symptoms, social history, and overall health status was not adequately provided. In other words, it was observed that developers found it difficult to include specific information in the scenario. The following are some of the points mentioned during the meetings by the medical education experts and SP trainer:

“In terms of social history, when the student asked the patient whether he or she drinks coffee, if the SP responds that he or she no longer drinks coffee, the student should follow up with a question to confirm how many cups the patient drank before quitting coffee. This isn't present.” (S1)

“Regarding the patient’s physical examination, please add more questions such as... ‘What motivated you to receive a health checkup?’ and ‘Is it just a regular check-up at your workplace, or did you receive a check-up because of abnormal symptoms?’ (S3)

“In the case of dizziness, in the question ‘Does the dizziness get worse if you overwork?,’ does overwork mean a mental or physical thing? Please explain more about it.” (S3)

3.3.5.2. Difficult for SPs to understand

Second, there was an opinion that the KM terminology and Chinese characters should be written in Korean so that the SPs can understand the scenario.

3.3.5.3. Concerns about differences between SPs within the same case

Third, in cases where several SPs demonstrate the same single case, it should be standardized as much as possible to ensure that there is no variation between their demonstrations. In this regard, the solution suggested to the developers was as follows. Because students may ask sudden questions while performing OSCE, it is necessary to inform the SP of the behavioral standardized guidelines, which cover
how to answer or act in response to a student's question. In addition, one more solution was presented to
the developers to clarify the timing of the question that must be asked. There were some questions from
SPs due to a lack of guidelines.

“Should I unwrap my watch to show my hand while the student is performing the pulse diagnosis? Or
should I unwrap only if the student tells me to take off my watch?” (S3)

“Should I present the examination card if only one of my hands is pulsed? Please tell me when or in
which situation I should present the physical examination card.” (S3)

“Should I withhold information about the medications I’m taking unless the student specifically asks
about that?”

Here is a suggested solution from a medical education expert:

“If you have a question for the SP, you should set up adequate time to ask it. If standardization is difficult,
it is preferable to eliminate the question entirely.” (S1)

3.3.5.4. SPs’ dialogue includes critical contents

Fourth, there were some problems where the SP gave some information to the student in advance, even
though the student did not ask them anything. So, according to the researchers, there is a need to develop
scenarios with certain guidelines, as above. In addition, a medical education expert suggested removing
questions such as “Doctor, aren’t you going to do a pulse diagnosis for me?” All developers modified their
scenarios based on the above solutions.

3.3.5.5. The range of patients’ education is limited

We argued about the purpose of this development, that is, developers should have in mind when making
a scenario. In previous projects about OSCE module development, there was a format that was
categorized as “patient education,” but in this project, that format was not used. So, two developers were
not sure if they should make a scenario about patient education in detail.

“In other CPG projects, they told the developers that the patient education item was missed, so they asked
us to develop [something] for that. I think that the ‘patient education’ item should be included because the
purpose of this project is to spread the CPG.” (D4)

In this regard, a medical education expert suggested the following:

“The part of ‘patient education’ should be developed focusing on following the diagnostic progress,
because the diagnosis wasn’t decided yet, so we couldn’t offer any kind of education to the patient.” (S1)

Another KM education expert suggested the following solution:
“In KM education the OSCE is the beginning stage, so we need to focus on not PPI but ‘OSCE performance’ like ‘history taking’ or ‘physical examination.’ So, when I reviewed your scenario, I removed the part on ‘patient education’ from my purpose.” (S2)

“Therefore, the patient education in the scenario mostly consisted of future diagnostic progress including life management, future plans, and so on.” (S2)

3.3.6. Format

There was some discussion on the need to use a standardized format because each developer used a different format to write the OSCE module. To resolve this, the researchers agreed to prepare a standard for matching the format in common through a meeting, as follows:

1. The present illness must be expressed in short sentences in a sequence so that students can read and memorize it more easily. As an example, 3 months ago, 1 year ago.

2. The format of the checklist should be expressed in the patient's words. For example, not “Did the student ask the patient when their symptoms started?” but “The patient said that he had been tired for about a year.”

3. It should be organized so that the contents are appropriate for each item. For example, in some cases, the contents of the physical examination were included in the history taking, so it should be revised. There were also issues with the contents not being organized consistently between the checklist and the scenario or the present illness and the case summary within the same scenario. For example, at the beginning of one OSCE module, it was shown that the patient was in his 60s, but at the end of the scenario, the patient was shown in his 50s. So, the developers were asked to correct these parts by the medical education experts.

3.3.7. Pattern identification

The issue most frequently raised by the developers was that they had to develop cases based on PI schemas. They became confused about whether the existing PI schema was well-crafted, so some of them even created their own PI schema for developing their OSCE modules. The following are some of the difficulties associated with PI.

3.3.7.1. The gap between schemas and real clinical situations

We categorized their difficulties by PI schema into three items. First, there was a significant gap between the inference based on the previously developed schema and the actual clinical situation. The developers wondered whether KM doctors were treating based on the existing schema in the actual treatment. They were also doctors, so they thought that many actual clinical situations would not follow the schema. As a result, they were not sure if it was allowable to create a OSCE based on the existing schema.
“The PI is really complicated. In my clinical situation, I mainly diagnose patients based on the stage of the disease not based on the schema. For example, in the case of stroke patients, if they are in the acute phase, I mainly diagnose them as ‘fire heat (火)’ or ‘strength (力)’ and, if the patients are in the chronic phase, they are mainly diagnosed as ‘Qi deficiency (气虚)’ or ‘Yin deficiency (阴虚).’ So, most clinicians generally diagnose based only on stage not schema, like me. In the case of a complex PI schema, can we consider the schema only by the stage of the disease?” (D1)

In regard to the above problem, a KM education expert replied that it does not matter if the contents of a OSCE differ from a clinical case. This is because the purpose of a OSCE is not to reflect the actual clinical field, but to train students to be familiar with schema-based deductive reasoning. Therefore, it needs to be developed based on schemas, and the developers tried to solve the issue based on the expert's opinion.

“Clinicians have a tendency for pattern recognition, which requires advanced training in inductive reasoning based on schemas. In other words, they consider three to five items at once for diagnosis and it leads to them making comprehensive decisions. However, in the case of students, their knowledge is dispersed, making it difficult to do pattern recognition. And as the goal of OSCE is education and evaluation for students, that's why you should focus on training students to become more like professors; that is, pattern recognition of experts, via schema-based reasoning training. So, even if the OSCE module is not similar to clinical practice, the goal should be for the student to imitate and practice the expert's treatment form, so schema-based development should be conducted.” (S2)

3.3.7.2. Uncertainty of the existing PI schema

The PI list was presented in the CPG. As the CPG is developed mostly based on clinical trials and has some differences compared to textbooks, the developers could not be sure that the PI list could be used for a model answer as-is. Furthermore, the listing of simple items does not reflect the clinical reasoning process, so a schema-type configuration was required as in the disease diagnosis. Therefore, there was a problem in determining which PI items to include and how to create the schema. The following topics were discussed at the meeting:

“The Conventional medicine schema is a type of systematic exclusion diagnosis that classifies and organizes possible diseases under a single symptom. On the other hand, the KM schema presented in the current CPG is not a type of exclusion diagnosis, but rather a parallel form that summarizes the possible PI diagnoses. I think these kinds of PI schema could not be considered as reliable schemas. So, I'm not sure if I should develop the module using the existing schema.” (D5)

“There may be enough challenges. Realistically, asking about all PI categories is impossible, and there are only a few key questions that we have to include. That's why we have no choice but to compromise the schema by considering various cases. I think that it would be better to create a new schema focusing on overlapping parts of the PI system in textbooks and the CPG.” (S2)
Accordingly, in our 2nd meeting, all researchers agreed that the PI schemas needed to be reconfigured, which was done by the developers during this project. After that, they continued to develop the following OSCE modules based on their new schemas. For this process, a KM education expert joined the study to provide appropriate solutions for schema reconfiguration. The following are the challenges encountered by the developers while reconstructing the PI schemas, as well as the solutions proposed by the KM education expert:

“Let’s reorganize the schemas by categorizing the related PI rather than dividing PI into ‘deficiency ( )’ and ‘excess ( )’. Because there are various diagnosis systems in KM, it is difficult to create totally perfect PI schemas. Let’s decide how we make the schemas by reaching a consensus among the researchers.” (S2)

3.3.7.3. Difficulty in selecting items for PI

The developers had many troubles in creating questions related to PI. Due to the characteristics of PI diagnoses, there can be many questions for only one diagnosis. For this challenge, it was suggested that the developers replace the PI-related questions with disease characteristic questions such as OLD CoEx CAFÉ, an abbreviation of onset, location, duration, course, experience, character, associated symptoms, and factor. In addition, it was suggested that rather than questions for selection, it should include questions for exclusion. All researchers agreed with the above solutions.

“For example, when I diagnose as spleen Qi deficiency syndrome in some cases, there are many PI-related questions in accord with it such as anorexia and stomachache. Should I include all these questions on PI-related symptoms? I’m not sure which of these questions I should ask in order to properly diagnose and assign a score.” (D4)

“Do not try to include all PI-related content as questions. In other words, kinds of questions about the characteristics of disease, such as OLD CoEx CAFÉ, could also be good questions for PI diagnoses. You might be relieved of the burden of making questions with this solution.” (S2)

Even though the developers solved the above issues based on the experts’ suggestions, they emphasized the need to develop a PI item list commonly applied to all diseases. In the PI diagnostic system, even if two patients have different diseases, the diagnoses can be the same. For this reason, the developers found some problems, as follows:

“Is there a significant difference between spleen Qi deficiency syndrome of indigestion and spleen Qi deficiency syndrome of chronic fatigue? In the case of gastrointestinal symptoms, for example anorexia, it can show both indigestion and chronic fatigue. I’m not sure how to tell them apart. It would be better to choose critical questions from a well-made PI item list.” (D3)

“The most difficult thing for me as a developer is that even though I am a clinical expert, I am concerned about my ability to create new PI schemas. I think that first, it should be set up with the goal of
distinguishing ‘deficiency syndrome’ and ‘excess syndrome’ at a higher level for differential diagnosis…”

(S2)

3.4. IPA Analysis

An IPA analysis was conducted to identify which challenges should be the focus when trying to solve the difficulties faced by developers. An IPA graph is divided into four quadrants with importance on the x-axis and performance on the y-axis (Figure 2).

According to the location of each item, four quadrants were named: “Possible overkill” (Quadrant 1), “Keep up the good work” (Quadrant 2), “Low priority” (Quadrant 3), and “Concentrate here” (Quadrant 4). “Possible overkill” is the area where developers’ performance compares well to importance and “Keep up the good work” is where the current level must be continuously maintained because both importance and performance are high. “Low priority” is an area that needs to be improved because both its importance and performance are low, while “Concentrate here” does not necessitate excessive effort due to its high performance in comparison to its importance.

In our results, the items included in the “Concentrate here” area were “PI schema reconfiguration” and “Making adequate PI-related items” under the PI theme and “Providing standardized guidelines” under the scenario theme. In the “Low priority” area, there were “Providing critical information,” “Adjusting the number of items,” and “Including critical items” under the test situation theme and “Prohibited telling critical items in advance by SP” and “Range of patients’ education” under the scenario theme. “Authentic case” and “Adjusting the number of items” under the checklist theme and “Providing sufficient information” under the scenario theme were found in the “Possible overkill” area. The “Keep up the good work” area included “Case selection,” “Setting scoring criteria and format,” “Marking items clearly,” “Using easy terms for SPs to understand,” “Observation format,” and “Consistent contents.”

Discussion

To the best of our knowledge, this is the first qualitative analysis of developers’ challenges and the related solutions encountered while developing OSCE modules. The developers related their experiences through meetings, surveys, and interviews, and adequate feedback from medical education specialists was obtained through the development process. Through the repeated OSCE development process, their self-confidence was increased. Thus, it was shown that the experience of OSCE development can help developers obtain the required skills for developing. However, they faced many difficulties before they reached that level.

Our main results were categorized into seven themes as the subjects of the developers’ challenges. The factors were further classified into 18 subcategories, some of which are discussed below. The first theme was about selecting adequate cases. The developers wondered which patient cases should be chosen for well-made modules. This challenge was handled by choosing common and usual patient cases rather than uncommon and special cases. Also, it was suggested that it would be better to diversify the patients’
gender and age for each case. This can lead students to consider varied cases. This challenge can be faced by developers at the beginning step of OSCE development. These solutions might be helpful for developers in further OSCE development.

Second, there were challenges centered on the test situation. The developers said that they felt very awkward when the SP was presenting the diagnosis card to the student in the test situation, because it was so far from a real clinical situation. In the method of suggesting a diagnosis card, when the student inspected the SP’s tongue or pulse, for example, the SP would present a card containing some KM examination results such as a yellowish tongue or weak or string-like pulse. The developers discussed how to present the examination results in the form of a photograph, but concluded that this would be difficult due to the issue of whether a photograph could be printed close to the actual color of the tongue. Finally, through presenting a minimum medical examination card, the challenge was resolved. The reason for this difficulty was that, in the case of medical education, not KM education, the method can be made clear by using a technique like auscultation or X-ray, so that the correct answers can be drawn for diagnosis, whereas it is difficult to standardize the KM diagnosis results in the case of the tongue and pulse [15]. As a matter of fact, despite tongue inspection being a non-invasive and simple method, it is not widely used. This is due to the fact that the result is affected not only by external factors such as the brightness of the light but also by subjective factors related to the experience, knowledge, and diagnostic skills of individual practitioners [16, 17]. In this regard, the diagnostic criteria for tongues are known to be unclear, so it is difficult to obtain consistent and reproducible results among practitioners [16, 17]. These issues have long been controversial within the KM community, and various attempts have been made to standardize KM diagnostic methods by many researchers [18, 19].

Third, the developers found it difficult to properly establish scoring criteria for the post-encounter note. In clinical clerkship education, deciding on reliable and reasonable scoring criteria is one of the most important things; however, most faculties find it difficult to do. In particular, the issue of assessment in clinical practice has been raised repeatedly. In one survey, faculty members pointed out a lack of valid assessment tools and criteria in practical education. In addition, Im et al. reported that to ensure the reliability and validity of OSCE assessments, interrater agreement should be secured. This is in accordance with our argument. Thus, the decision on scoring criteria is not an individual problem for developers, and efforts by institutions and evaluators to set up consistent and valid criteria are required.

The fourth factor centered on the checklist. The developers had much trouble deciding on the number of questions. Since the OSCE is about the KM field, it is supposed to include KM diagnoses as well as Conventional medical diagnoses, so excessive items have been included. The developers were not sure how many PI-related items should be included. When an KM clinician makes a diagnosis, they integrate objective diagnostic indicators like facial color, body shape, tongue condition, and pulse with information about the patient’s subjective symptoms like digestion, shape or color of stool and urine, temperature sensation, and sleep state. That means that even though only items related to PI were included in the module, there was still an excessive number of items. One study reported that in cases of stroke, 122 PI indicators were obtained, even though common symptoms were excluded [20]. Thus, excessive items
were found in our study even though the developers had tried to include only critical items related to the subject. Regarding this problem, all researchers concluded that PI-related items must be included in order to distinguish KM from medical education, since KM diagnosis is the most important part of KM education. This issue is also related to the problem of PI theme, so finally all researchers decided to reconstitute the PI schemas with the help of a KM education expert.

The greatest concern of the developers was problems in implementing PI. PI is a unique diagnostic system that integrates the patient's symptoms, the origin and location of the disease, and nature, and also physical examination methods such as inspection, olfaction and auscultation, inquiry, and palpation [21]. Even when two patients are diagnosed differently by conventional medicine, the results of the PI diagnoses might be the same. These characteristics have advantages, such as one-to-one personalized care; however, the critical problems of non-standardization and ambiguity remain. According to one previous study, when KM clinicians diagnose patients, the most challenging part is “the lack of objective diagnostic indicators and standardized PI. [10]” In our findings, it was obvious that the difficulties caused by the lack of objective indicators and the non-standardized characteristics of KM were the most significant issues for the developers as well. They solved these problems by creating new PI schemas using the schema reasoning method. The new PI schemas were created by grouping similar PI indicators and categorizing them together, as well as determining the level of stratification based on the disease's severity. In this process, the developers said that it was very helpful to develop the new schemas with the assistance of a KM education expert. The expert's opinion was as follows. Decisions at each branching point in the diagnostic schema are validated by key predictors (signs and symptoms), and one subcategory can be selected while excluding others, leading to an efficient and accurate diagnosis. In the same manner, if the structure of a PI schema is redesigned based on the characteristics of the PI types and mechanisms derived from the experience of the expert or scientific evidence, the PI can be adopted or excluded at the branch points, so that a focused and efficient PI can be achieved; thus, checklists can be written efficiently. Additionally, they emphasized that a further OSCE development project should begin after developing a well-made schema with a standardized diagnostic system. That is, we could see how difficult it was for the developers to create PI schemas. The results from the IPA analysis also testified to the above difficulties, as “PI schema reconfiguration” and “making adequate PI-related items” were included in the “Concentrate here” area. This shows that the standardization of PI diagnosis is important not only for diagnosis, but also for schema development in the future for the sake of curriculum development.

In our IPA analysis, “making PI-related items,” “PI schema reconfiguration,” and “providing standardized guidelines” were the items that the developers considered to need the most improvement. Therefore, the efforts of the KM community are needed to ensure the standardization of PI diagnosis and PI-related item indicators.

Interestingly, the item “too many items” was found to be in the “Low priority” area, even though the developers complained very often about this issue. This is most likely due to the developers considering it easier to change the number of items because they thought all checklists should be included. However,
this result is thought to be due to a lack of insight by developers who only participated in the OSCE development stage. One previous study reported that creating a new hybrid item by combining each existing item to reduce the number of items could not only solve the problem of time limitations, but also improve the reliability of the medical practical test [22]. Thus, this should be thought of as a critical issue that must be resolved during the development process. To enhance developers’ insight, we think that giving them the experience of performing the entire OSCE process from training SPs to scoring students’ performance would help to improve their awareness of the importance of each item. In addition, even though the item “missing critical differential diagnosis” was mainly included under the checklist theme, the developers also recognized it as being of low importance. This seems to be a result of the developers thinking that other items were less important because they primarily focused on PI issues. Another possible reason is that the developers’ perspectives on whether it is a critical question or not differed based on their specialization and clinical experience. Previous research on the development of PI indicators reported that clinicians of the same specialization could potentially conduct different diagnoses in a clinical situation. Also, clinicians who are already accustomed to illness scripts might not consider the need to make the items according to the schema in their thinking system. These are considered to be the most difficult things to improve in the process of developing a OSCE module, as both importance and performance were recognized as low.

Furthermore, in our interviews, the developers mentioned deficiencies and problems in the educational conditions. They reported that sufficient space, facilities, and human resources are needed to conduct effective OSCE education. They especially emphasized the need for a system with medical education experts and SP trainers to guide and train faculty members. Even though the developers enhanced their competency through feedback from medical experts and an SP trainer, it is suggested that institutional support is needed to improve OSCE performance.

Our study has a few limitations. First, it was conducted as part of a single project about an internal medicine OSCE, so the generalizability of the results may be limited. Another limitation is that we solved the challenges with help from only two medical education experts, so various other solutions could be possible.

As far as we know, this is the first qualitative analysis to investigate the experience of developers during OSCE module development for KM education. This experience improved the developers’ personal competence in developing OSCE modules, and we identified several issues and related solutions that they encountered during this developing project. Our hope is that this study can help guide developers in further OSCE development projects. Although the fact that the developers’ competence increased is a very positive aspect, the most difficult part for the developers during this project was the ambiguity of the PI diagnosis and the existing schemas. This suggests that to develop the KM clerkship curriculum there is an urgent need for support from the oriental medicine education community and schools, and we need to strengthen individual developers’ capabilities.

**Declarations**
Ethics approval and consent to participate

All developers took part voluntarily and agreed to the study process. Ethical approval was obtained from Gachon University Institutional Review Board (GRIB-21-108). Written informed consent was obtained from all participants. All methods were performed in accordance with the 'Declaration of Helsinki'.

Consent to publish

Not applicable

Availability of data and materials

The datasets used during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by a grant from the Guideline Center for Korean Medicine, National Institute for Korean Medicine Development (HI16C0275) & Biomedical Research Institute Grant (202200370001), Pusan National University Hospital.

Acknowledgement

Not applicable

Authors’ contributions

ARJ and HYL prepared the first draft of the manuscript and collected and analyzed the data. SI supervised the study and made substantial contributions to its concept and design. SWS and SI participated in the analysis and interpretation of the data, and proofread the manuscript. Finally, all authors have read and approved the final manuscript.

References


22. Han JJ. A study on the improvement of question types and test methods in the doctor’s national exam. Korean Health Personnel licensing Examination Institute. 2013. [Korean article]

Figures

![Graph showing self-assessment of confidence in developing OSCE cases](image)

**Figure 1**

Self-assessment of confidence in developing OSCE cases
Figure 2

The importance-performance analysis of major challenges

1, Case selection; 2, Authentic case; 3, Providing critical information; 4, Setting scoring criteria and format; 5, Adjusting the number of items; 6, Including critical items; 7, Marking items clearly; 8, Adjusting the order of items; 9, Providing sufficient information; 10, Using easy terms for standardized patients to understand; 11, Providing standardized guidelines; 12, Prohibit standardized patients from talking about critical items in advance; 13, Range of patients’ education; 14, Observation format; 15, Consistent contents; 16, Making items based on structural schema; 17, pattern identification schema reconfiguration; 18, Making adequate pattern identification-related items

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.