Medical Teacher

The use of a virtual patient case in an OSCE-based exam - A pilot study

O. Courteille a; R. Bergin†a; D. Stockeld a; S. Ponzer a; U. Fors a

a Karolinska Institutet, Stockholm, Sweden

First Published: 2008


To link to this article: DOI: 10.1080/01421590801910216
URL: http://dx.doi.org/10.1080/01421590801910216

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
WEB PAPER

The use of a virtual patient case in an OSCE-based exam – A pilot study

O. COURTEILLE, R. BERGIN, D. STOCKELD, S. PONZER & U. FORS
Karolinska Institutet, Stockholm, Sweden

Abstract

Background: This study focuses on a skills test based clinical assessment where 118 fourth-year medical students at the four teaching hospitals of Karolinska Institutet participated in the same 12-module OSCE. The goal of one of the twelve examination modules was to assess the students' skills and ability to solve a virtual patient (VP) case (the ISP system), which included medical history taking, lab tests, physical examinations and suggestion of a preliminary diagnosis.

Aims: The primary aim of this study was to evaluate the potential of a VP as a possible tool for assessment of clinical reasoning and problem solving ability among medical students. The feeling of realism of the VP and its possible affective impact on the student’s confidence were also investigated.

Method: We observed and analysed students’ reactions, engagement and performance (activity log files) during their interactive sessions with the simulation. An individual human assistant was provided along with the computer simulation and the videotaped interaction student/assistant was then analysed in detail and related to the students’ outcomes.

Results: The results indicate possible advantages of using ISP-like systems for assessment. The VP was for instance able to reliably differentiate between students’ performances but some weaknesses were also identified, like a confounding influence on students’ outcomes by the assistants used. Significant differences, affecting the results, were found between the students in their degree of affective response towards the system as well as the perceived usefulness of assistance.

Conclusion: Students need to be trained beforehand in mastering the assessment tool. Rating compliance needs to be targeted before VP-based systems like ISP can be used in exams and if such systems would be used in high-stake exams, the use of human assistants should be limited and scoring rubrics validated (and preferably automated).

Introduction

Over recent years, we have seen an increasing level of use of simulated and virtual patients (computer-based simulations of patients) for both training and assessment in medical education (Cantillon et al. 2004; Guagnano et al. 2002). This has been extensively reported by Issenberg et al. (2005) in a BEME systematic review (guide no 4) where both problems and opportunities were thoroughly investigated.

Studies investigating the usefulness of simulations for instance, have showed that students might react similarly to real and simulated patients (Sanson-Fisher & Poole 1980). In the study by Edelstein et al. (2000), students thought that computer-based case simulations were better tests of clinical decision making than written shelf examinations. Schuurh & Van der Vleuten (2003) investigated the problem of construct and face validity and described how to meet the need for assessment procedures that are both authentic and well-structured.

In their educational assessment guidelines, Appel et al. (2002) even recommend the Clerkship Directors to ‘use computer-based case simulations to augment traditional internal medicine evaluation methods’ that these ‘would be used as a supplement to, and not a replacement for, other assessment tools’ (http://dx.doi.org/10.1016/S0002-9343(02)01211-1). In a special themed article, Holmboe (2004) states that simulated patients and other simulation technologies are considered as ‘being important and reliable tools for teaching clinical skills and evaluating competence’ but also emphasises that ‘they cannot substitute to the direct observation by faculty of trainees’ clinical skills with actual patients’. Hence, because of deficiencies in faculty direct observation evaluation skills, automated scoring of patient interactions has been proposed as a way to limit the effort required for mentor evaluation (Nielson et al. 2003). However, we need to be aware

Practice points

- Virtual Patient cases with automated scoring might be used as a complementary method for summative assessment.
- Students need to be trained in mastering the assessment system prior to exams.
- Scoring rubrics should be developed and validated before implementing computer-based assessment.
- Human assistance should be limited because of possible confounding influence on students’ outcomes.

Correspondence: Olivier Courteille, Karolinska Institutet, Berzelius väg 5, S-171 77 Stockholm, Sweden. Tel: 46-8-524 8 7269; fax: 46-8-34 51 28; email: olivier.courteille@ki.se

†Deceased.
of the necessity first to ensure the construct validity of automated scores (Collins & Harden 1998).

Simulated patients have therefore been suggested to be useful as assessment tools in Objective Structured Clinical Examinations (OSCEs) or in other assessment methods in evaluating students’ interactions with patient related medical issues, such as clinical reasoning and/or medical problem solving abilities (Collins 1992; Schuwirth & Van der Vleuten 2003). More recently, virtual patients (VP) have been gradually introduced as a complementary method to simulated patients because they support active and reflective learning (Clyman & Orr 1990; McGaghie 1999).

Interactive simulation of patients (ISP)

ISP is a comprehensive and high-fidelity virtual patient-based learning tool designed for medical and healthcare students (clinical level) to explore and solve clinical cases in respect of diagnosis (Bergin & Fors 2003). The system aims to help students to practise clinical reasoning skills and has been designed for that purpose to resemble a realistic patient encounter covering extensive functions for medical history, physiological examination procedures and laboratory/imaging tests.

To achieve a sense of authenticity in a virtual healthcare situation, the ISP enables students to freely navigate without any pinpoints and choose, for example, to:

- take the patient’s illness history and ask any medical related question (via an interactive dialogue with text entry, natural language processing and video-clip based answers), (see Figure 1).
- ‘perform’ any physical examination (most medical examination procedures are available); (see Figure 2).
- request one or several laboratory/imaging tests,
- suggest a preliminary diagnosis,
- go back to either the illness history section, or order more lab tests and/or perform physical examinations,
- ask for feedback.

The ISP is also designed to act as naturally as possible, meaning that the virtual patient can be programmed to ‘react’ emotionally on, for example, repeated questions, unnecessary questions, or unwanted questions related to sexual behaviour (if not medically indicated) (Bergin & Fors 2003; Bergin et al. 2003).

Although the case scenarios are pre-defined, the interaction itself is unscripted. The scenario establishes initial conditions but the student’s responses to the virtual patient, as well as...

Figure 1. ISP screenshot from the history taking. Note the free text input in natural language.
inherent flexibility in how the ISP patient is allowed to react, aim to result in an authentic case interaction. ISP is also capable of tracking the individual learner’s decision paths and interactions during the case encounter in detail.

Previous studies have shown that ISP is an engaging and trustworthy learning tool (Bergin et al. 2003) with potential to offer realistic patient case scenarios, therefore, it has been suggested to use this case simulation method as one of several possible assessment methods in a clinical exam.

Objective structured clinical examination

OSCE is a well-known assessment method that has been developed for use in both formative and summative assessment of students’ clinically related knowledge and performance (Harden 1990; Clyman & Orr 1990). The basic ideas behind the OSCEs are that this type of assessment is structured – trying to be more objective than the merely passive observation during clinical rotations, and that they are intended to assess clinically relevant procedures like patient interviews, medical decision making and practical tasks (Harden et al. 1992; Newble 1992).

However, OSCE exams also have some potential problems such as that they can be very resource intensive in terms of the number of teachers/observers needed as well as the logistics involved in terms of time allocation and facilities (Van der Vleuten et al. 1989). Even though they are intended to be structured and objective, the judgments are still made by individual teachers/evaluators whose assessments criteria might jeopardize both reliability and validity (Weatherall 1991; Wilson et al. 1969).

Aim

The primary aim of this pilot study was to assess the potential of a VP case as part of an examination of clinical reasoning and problem solving ability among medical students. Our research questions were the following:

- Can a VP learning tool like ISP be used for assessment and thus be able to differentiate reliably between students’ performances, including gender-related differences?
- What are the necessary modifications and do the students need to be trained beforehand?

Secondary aims were to try to measure the feeling of realism of the ISP system and perform a preliminary observation of the affective impact of the VP on the student’s confidence and ability to solve clinical problems. This led us to formulate an additional research question as follows:

- Can we notice an emotional incidence on the social interaction despite time and situational constraints?
Methods

The OSCE exam

The study was conducted in May 2004 during the OSCE procedure at the surgery course of the five and a half-year long medical programme at Karolinska Institutet (KI), Stockholm, Sweden. The general examination goal of the OSCE in the surgery course was to assess the students’ skills and ability to perform basic physical examination (e.g. examination of the knee joint) or procedures (e.g. local anaesthesia) and also to assess the students’ ability to carry out a systematic patient interview, in order to figure out the diagnosis and suggest adequate examination and treatment as defined in the curriculum for the course. During the examination, the students rotated through a series of 12 different skills stations during 90 minutes. They had to perform a specific task at each station and their performance was scored by an observer using a predetermined checklist or rating scale. Different session lengths were allocated to the stations depending on the task: 5 minutes for each of the eight ‘short stations’ and 10 minutes for each of the four ‘long stations’. At each hospital, four of the long stations consisted of two ISP’s and two Standardized Patients (SP) who were human actors, see Figure 3. The common goal for the participants at both the SP and ISP stations was to take a short history and then either to inform the patient about the findings (SP) or to make a diagnosis (ISP) within 10 minutes.

All of the fourth year students (n=118), enrolled in the 20-week clinical course in surgery participated in the OSCE exam. They were dispatched over four university hospitals (later on referred to as H1, H2, H3 and H4) and divided in two subgroups of up to 16 students each.

The ISP station and the case

Since ISP was not originally designed for assessment purposes, special customizations and set-ups had to be considered. Two surgeons were involved in the redesign process of an existing colorectal case (requiring about 16 person-hours in total). The customized technical design and content management was performed by two developers and required about 48 person-hours.

Furthermore, none of the students had tried the ISP system before, so in order to facilitate the use of ISP in this pilot study, every ISP station was assigned an assistant who knew how to run the system. Moreover, since this was a pilot test, the performance on the ISP station was decided to only be able to result in positive outcomes for the students, meaning that their overall test score might increase if the ISP case was handled very well, but that no student could be down-graded due to a bad result on ISP. Thus no student would fail the OSCE exam because of a bad performance on ISP.

The basic requirements for solving the ISP case were considerably simplified as compared with the cases used for learning, in order to make it possible to complete the case
within 10 minutes. For instance, there were fewer lab tests available in the simulation than usual (but still enough for not providing too many clues and/or revealing the nature of the case) and the illness history was made unusually straightforward with a very ‘cooperative’ patient. The case used was based on a 68-year-old female with rectal cancer and with a relatively clear medical history and symptoms.

To facilitate grading, the case creator (an experienced clinical teacher with expertise in the actual medical domain) was asked to specify the most important illness history questions as well as the most relevant physical examination procedures and laboratory tests. Two other senior clinicians also validated the case. In total, 27 illness history questions (out of which 11 were judged as required), 12 physical examinations (out of which 10 required) and 11 lab tests (out of which 7 required) were identified as important for this ISP case.

When running the ISP case, none of the required questions/procedures or any feedback on these were revealed to the students, until a correct diagnosis had been submitted. The system was also automatically recording which interactions were made with the system. The results were stored in a database format for enabling further analyses, for instance, computing how many of the most relevant questions had been asked and procedures undertaken, in what order they appeared and the time elapsed for each task.

**Assistants**

An assistant whose task was to introduce the system briefly for the student and help him or her to navigate and interact smoothly with the programme was available at each ISP station (n = 8). The assistant was also instructed, if needed, to help the student to formulate system compliant history questions by offering to handle the keyboard and/or get the student back on track if he or she was deviating too much from the objectives (due to lack of familiarity).

In order to avoid interference with the results, the assistants should not reveal the correct diagnosis for the students or supply too much ‘medical help’ in solving the case. When the ISP session was over, the assistants handed out a paper-based summary of the case, including formative feedback to the examinees.

The enrolled assistants were physicians, clinical teachers or other persons who were very familiar with the ISP system.

**Data collection**

The ISP system automatically tracked each student’s interaction. This information, gathered in log files, contained complete and detailed chronological data on medical history questions asked, physical examinations performed and lab tests ordered, interaction time and navigation paths, and finally diagnoses suggested/submitted, confidence scale and justification of the suggested diagnoses.

Two questionnaires were also conducted during the OSCE procedure. A main questionnaire (online-based and delivered as one of the four long stations) evaluating the surgical course and the OSCE as a whole, and a second questionnaire specific to the ISP (paper-based and handed out to the participants immediately after the ISP sessions). The students were asked to return the ISP-specific questionnaire by post. This questionnaire was anonymous and collected information on demographic data as well as on usability, attitudes and expectations towards the ISP as a new examination tool.

**Video observations**

Additionally, in order to measure the possible affective impact of ISP on examinees, video observations were performed on four of the eight ISP stations. A DV-camera was placed behind the computer screen and used a wide angle for covering both the student’s and the assistant’s mutual interaction, as well as the student’s own interaction with the ISP system. A coding process for further analysis was developed and based on the following variables: interaction assistant/student (weak, medium, intensive); expression of uncertainty (doubtful, neutral, certain); external signs of stress (low, medium, high); flow (frustrating, normal, playful); mouse handling (student, assistant) and finally keyboard handling (student, assistant).

An exact binary logistic regression analysis (LogXact 7.0, Cytel Software Inc.) was carried out to investigate the association between indicators like the behavioural variables described above (exposure variables) and the student’s outcomes (outcome variable).

**Grading/assessment potential**

To study the potential of ISP as an assessment tool to differentiate between students’ individual performances, an initial hypothesis was set up: there should be ‘enough’ variability among the students’ individual results in solving the case (Friedman Ben-David 2000). Typical ordinary exams at KI usually have a level for passing set at about 70–80% correct answers, and the results of most exams show that most students pass this level, but that some are well below that level on the first trial. Therefore, we set a ‘pass rate’ of about 70–80% on the ISP station as a goal.

To study the potential of ISP to be able to present cases with accurate complexity, the variability in suggesting a correct diagnosis, the percentage of correct illness questions, physical examinations and lab tests proposed were also measured. Furthermore, to judge this pilot test as positive, it was stated that most students should be able to come to a preliminary diagnosis within the time limit allocated to the long OSCE stations (10 min). The completion time was therefore computed.

**Ethical considerations**

This study was approved by the ethical committee of Karolinska Institutet. In order to comply with these ethical considerations, a consent form was handed beforehand to every examinee informing him/her about the specific aspect of the ISP station, including possible video observation, and that the ISP station could not influence their grading in a negative way.
Overall results

Altogether, 110 students out of the 118 in the course participated in the OSCE examination. The eight missing were either sick or had other allowed excuses. All 110 students volunteered to use the ISP station. The overall performance can be observed in Table 1 below.

As can be seen, the 110 students worked with the ISP case for about 7 minutes 45 seconds on average. Most of the students arrived at a correct diagnosis after 1.12 tries on average. However, as indicated in Table 1, the assistants at hospital 1 seemed to have helped the students to arrive at a correct diagnosis, since a larger than anticipated majority of the students at hospital 1 found the correct diagnosis. These two assistants later on confirmed this assumption in part. Furthermore, the students at hospital 3 had only 40% (group 1) and 42% (group 2) correct diagnoses on average, which might indicate that their assistants were less helpful than the other assistants. This was also indicated by the fact that the same students ran out of time, limiting the possibility of submitting a second diagnosis. However, it is interesting to note that all students who had time to submit a diagnosis already succeeded on the first trial. Figure 4 shows the direct effect of the degree of assistance on the mean session time.

In Table 2, the detailed results from the medical history and lab sections of ISP are shown. Unfortunately due to an unanticipated error in the logging system, the physical exam procedures were not logged. As can be seen, on average 8.8 questions were asked to the patient and 63% of these were highly relevant. Those who asked the most history questions (e.g. hospital groups 2.1, 3.1 and 3.2) formulated them with a higher number of words per sentence. As a result, their completion time (above 8 min) was also longer than for other groups. On average,
5.6 lab tests were ordered, out of which 51% had been recommended by the case author.

The gender-related performance can be observed in Table 1 and Table 2. We observed that female students had generally higher means of required history questions asked (69% compared to 57% for males, \( p = 0.006 \)) and lab–tests ordered. Their overall performance was also slightly better compared to male students.

**Assistants**

Most of the students reported that the assistants helped them to feel more calm and comfortable. As a matter of fact, their presence was experienced as a relief for first time users’ in particular for the more stressed students. It turned out that this somewhat special (and resource-consuming) pilot test environment provided a convenient method to monitor the students’ clinical reasoning processes in a natural way. From analyses of the videotapes and from discussions with the assistants, it was found that the presence of the assistant seems to have made the students think aloud spontaneously and to verbalize what they were doing and why they were doing it while working with the ISP system.

Nevertheless, as reported later on, the assistants clearly influenced the performance of the students.

**Questionnaires**

The ISP-related questionnaire was answered by 68 students out of 110 (62%). Their median age was 26 years. The results of the questionnaire are shown in Table 3. As one could expect, due to the short time allocated to the ISP station and the fact that the students had never used the ISP system before, not all students’ answers were positive. The major reported complaint (question 2b) was about the limitations of the interactive dialogue with the patient. This might be explained by the fact that none of the students had used the system before, and to a certain degree also by the fact that the case’s dialogue interaction had not been fine-tuned.

However, the majority of students expressed agreement with the potential of ISP-like systems in future exams. Most of the students reported that they experienced the VP case as engaging (65%) and realistic (78%).

In Table 4, the ISP-related questions of the general on-line delivered questionnaire are shown. All students answered this questionnaire. However, due to the rotation scheme of the stations, 16 students (e.g. four students at each hospital) filled it out before they had used ISP (their blanked ISP-related answers were not considered). As can be seen, the opinions of the students differed to some extent from hospital to hospital. The overall opinions about the surgery course and the OSCE exam were rated higher than the ISP programme. This can be observed for students from hospitals H3 & H4, whose relatively bad performance and/or quality of assistance can be associated with rather low ratings for ISP. The very short and first-time experience with the ISP might also have conveyed a negative or diffuse general impression.

**Video observations**

Due to limited human resources we could only video-monitor half of the ISP stations. As a result, 47 students out of 110 were videotaped during their ISP session. No apparent effect from the presence of the video camera could be noticed on students’ performance. In fact, one female student reported that she ‘didn’t feel as nervous as being filmed during an encounter with a ‘real’ patient’.

As hypothesised, the quality and intensity of assistance provided had a strong positive impact on students’ outcomes (Table 5). The logistic regression analysis showed that the effect of a high degree of interaction between assistant and student (e.g. an intensive assistance) provided an estimated odds ratio of 17.21 (95% C.I. [1.30; 1032], \( p = 0.025 \)) when compared to a weak interaction (baseline group). This indicates that there is a significant correlation between the student’s outcome and the interaction between the student and the assistant.

The statistical analysis also showed that the assistant cluster 3 (senior physicians) and assistant cluster 2 (researchers experienced in medical simulations) had odds ratios of 17.37 (95% C.I. [2.48; 763], \( p<0.001 \)) and 2.77 (95% C.I. [0.84; 10.82], \( p=0.105 \)) respectively when compared to the (less experienced) assistant cluster 1 (undergraduate students of

<table>
<thead>
<tr>
<th>Hospital group</th>
<th>( n )</th>
<th>Mean no. of history questions asked</th>
<th>Mean % of history questions required</th>
<th>Mean no. of words per question</th>
<th>Mean no. of lab tests ordered</th>
<th>Mean % of lab tests required</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>17</td>
<td>6.9</td>
<td>78%</td>
<td>1.6</td>
<td>3.2</td>
<td>73%</td>
</tr>
<tr>
<td>1.2</td>
<td>12</td>
<td>7.8</td>
<td>77%</td>
<td>3.4</td>
<td>5.8</td>
<td>56%</td>
</tr>
<tr>
<td>2.1</td>
<td>11</td>
<td>10.2</td>
<td>61%</td>
<td>4.6</td>
<td>5.4</td>
<td>48%</td>
</tr>
<tr>
<td>2.2</td>
<td>15</td>
<td>9.1</td>
<td>62%</td>
<td>2.8</td>
<td>6.9</td>
<td>44%</td>
</tr>
<tr>
<td>3.1</td>
<td>15</td>
<td>10.0</td>
<td>60%</td>
<td>4.7</td>
<td>6.7</td>
<td>53%</td>
</tr>
<tr>
<td>3.2</td>
<td>12</td>
<td>9.9</td>
<td>56%</td>
<td>3.8</td>
<td>5.9</td>
<td>32%</td>
</tr>
<tr>
<td>4.1</td>
<td>12</td>
<td>7.7</td>
<td>50%</td>
<td>3.6</td>
<td>5.9</td>
<td>38%</td>
</tr>
<tr>
<td>4.2</td>
<td>16</td>
<td>8.7</td>
<td>62%</td>
<td>3.5</td>
<td>5.1</td>
<td>50%</td>
</tr>
<tr>
<td>Males</td>
<td>36</td>
<td>8.4</td>
<td>57%</td>
<td>3.7</td>
<td>5.8</td>
<td>46%</td>
</tr>
<tr>
<td>Females</td>
<td>59</td>
<td>8.5</td>
<td>69%</td>
<td>3.3</td>
<td>5.1</td>
<td>55%</td>
</tr>
<tr>
<td>Totals/averages</td>
<td>8.8</td>
<td>63%</td>
<td>3.5</td>
<td>5.6</td>
<td>51%</td>
<td></td>
</tr>
<tr>
<td>( P)-value</td>
<td>0.606</td>
<td>0.006</td>
<td>0.166</td>
<td>0.297</td>
<td>0.203</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Mean values (amounts and percentages) for history questions asked and lab tests ordered.
### Table 3: Median and average values for the ISP-related questionnaire.

<table>
<thead>
<tr>
<th>Question</th>
<th>Median</th>
<th>68 responses: 45 females, 21 males (2 did not report gender)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Was the ISP designed in such a way that you could apply your knowledge?</td>
<td>Median: 4 (scale ranging from 1 “Highly disagree” to 6 “Highly agree”)</td>
<td>52 responses</td>
</tr>
<tr>
<td>2a. What do you think is the best thing about the ISP examination?</td>
<td></td>
<td>52 responses</td>
</tr>
<tr>
<td>1. Realism/authenticity/trustworthiness of the case:</td>
<td>10 (15%)</td>
<td></td>
</tr>
<tr>
<td>2. Fun/enjoyable/engaging:</td>
<td>9 (13%)</td>
<td></td>
</tr>
<tr>
<td>3. Easy access to lab tests and physical exam:</td>
<td>8 (12%)</td>
<td></td>
</tr>
<tr>
<td>4. New learning mode/instructive/educational:</td>
<td>7 (10%)</td>
<td></td>
</tr>
<tr>
<td>2b. What do you think is the worst thing about the ISP examination?</td>
<td></td>
<td>64 responses</td>
</tr>
<tr>
<td>1. Limitation of interactive dialogue with patient:</td>
<td>38 (56%)</td>
<td></td>
</tr>
<tr>
<td>2. Lack of time:</td>
<td>14 (21%)</td>
<td></td>
</tr>
<tr>
<td>3. Doesn’t feel authentic:</td>
<td>10 (15%)</td>
<td></td>
</tr>
<tr>
<td>4. Not familiar with the system/this way of solving problem:</td>
<td>5 (7%)</td>
<td></td>
</tr>
<tr>
<td>3. Should stations of ISP type be used in the practical assessment test (OSCE)?</td>
<td>Yes</td>
<td>24 (35%)</td>
</tr>
<tr>
<td>4. What made the ISP case engaging?</td>
<td>Pos</td>
<td>42 (63%)</td>
</tr>
<tr>
<td>5. What contributed to some sense of realism in the ISP case?</td>
<td>Pos</td>
<td>53 (78%)</td>
</tr>
</tbody>
</table>

**Examples of positive answers to Q4:**
- Good structure. With history taking, status etc.
- Good to see the patient and hear her voice.
- To be able to think and reason with the help of the computer. Good complement to other teaching.
- The many possibilities to examine a patient. Great variation. Interesting to be able to try many tools as in real life and see what they would give you.
- The feeling of getting it right – the detective work – and having access to all options in an immediate way.
- The motivation to ‘solve’ the case.
- To be able to have examination done immediately and also to get results right away!
- Exciting, fun to get results from lab directly, good exercise. That the patient was able to answer my questions.
- Direct feedback, often missing in real life.
- To be able to freely choose examinations/tests.
- That the patient describes her symptoms in her own way – not just in textbook fashion.

**Examples of negative answers to Q4:**
- I became mostly irritated on the case.
- Nice picture, and the way examinations were done, but I don’t think it was efficient. Good pictures and examinations though.
- I was fun to try to reach the right diagnosis until I realized that she never would answer fundamental, important questions and that she began saying stuff that I already had asked without properly answering the question. That only made me irritated and I went on to do a physical examination instead.
- That I wanted to find out the problem of the patient. But it didn’t go smoothly all the time.
- Too stressful for an examination situation to become engaged. Under other circumstances can I imagine that one would become engaged because it was rather realistic.

**Examples of positive answers to Q5:**
- That the lady talks to me.
- To see the patient and that she talked to you.
- The voice and the pictures.
- The ‘living’ patient.
- It felt interactive, it affected the outcome.
- Real patient on the screen, with a voice. The variation of diagnostic possibilities.
- That the patient talked and moved. Not just replied with text.

**Examples of negative answers to Q5:**
- I think the interaction with patient didn’t work, otherwise interesting.
- A real person speaking, unfortunately she always said the same things.
- It didn’t feel realistic, unfortunately.
- The difficult history taking.
Medical Informatics), meaning that differences existed among the students in the perceived usefulness of assistance. The student's outcome also appeared to be strongly associated to the flow experienced with the virtual patient (i.e. the degree of affective response).

Grading possibilities

As mentioned above, no students could be 'down-graded' due to their performance on the ISP-station in this pilot study, but they could benefit from a good outcome on the ISP station as a way to pass the whole OSCE-exam (if they were short of only one or two points in the other stations). Therefore, the assistants were asked to fill in a special form for each of the students, indicating their overall performance on the ISP station.

In analysing these forms, it was rather clear that most assistants indicated a potential of using the ISP station for grading. Items like students' individual behaviour, ability to formulate adequate history questions, flexibility in re-formulating initial diagnostic strategies, combined with analyses of the individual log files, were mentioned as positive opportunities.

As observed above, the inter-hospital differences were higher than anticipated, with the inter-standard variability of important illness history questions asked (averages of 63%: S.D. = 18.99), of required labs ordered (average of 51%: S.D. = 26.60) and of correct diagnoses (average of 74%). This indicates, as hypothesised, that there was a rather large variability among the students' ability to solve the case, thus indicating that ISP-like systems might be used as one part of an 'assessment toolkit' for assessing students abilities to solve clinical cases. Even though the range of correctly performed physical examination procedures was not recorded, the assistants indicated that there was a satisfactory variability also in this aspect. The proportion of students supplying the correct diagnosis on first trial was 62%.

Discussion

This pilot study investigated the potential of a VP-based system tested during an OSCE-exam. Even though there are a number of methodological shortcomings and limitations, interesting results were found.

The case used and the assessment results

Although the case presented at the ISP station was re-designed to be easy to solve, it turned out that, given the time constraints, it was difficult for some examinees to solve the case without any external help. However there were clear indications that, if students can be trained and run a mock exam first, and also be given some more time to solve the...
cases, ISP-like systems might have a potential as a summative assessment tool (i.e. without human assistants). Earlier studies of ISP indicated that the system is fairly easy to learn, requiring only 20 minutes or so to learn how to interact with the system (Bergin & Fors 2003).

It can be noted that ISP was originally specially designed for collaborative learning, and previous studies have showed that the most ideal situation for solving simulated clinical problems is peer-to-peer collaboration. However, during the OSCE exam test here, the students could only rely on the assistant’s ability and willingness to provide medical advice, even if the latter were told to act as neutrally as possible.

This unique situation fostered the thinking aloud process on the examinee’s part. Consequently, it allowed the assistant to cope with the student’s ongoing clinical reasoning process and thereby to canalize his/her preliminary thoughts in case he/she was deviating from the main track. The observational data collection from the 47 video recordings on the ISP stations showed clearly that many of the examinees were unconsciously thinking aloud while an observer/rater was sitting beside them. The quality of the interaction and the degree of engagement from the assistant also appeared to affect the overall performance. This is something that needs to be further studied.

Variation in assistant behaviour and intervention

In their recent book, *Developing Organizational Simulations: A Guide for Practitioners and Students*, Thornton & Mueller-Hanson (2004) emphasize the importance of using ‘…trained assessors to observe behaviour, classify behaviour into the dimensions being assessed, and make judgments about participants’ level of proficiency on each dimension being assessed’ (Thornton & Mueller-Hanson 2004, p. 5).

The assistants at the ISP stations could act both as instructors and as raters, even if they did not grade the students in detail in this pilot study. But we do not know how many students actually worked on the case independently (i.e. with almost no help from the assistant). We identified problems with scoring validity due to the fact that the assistants/examiners’ level of help was not standardized, thus compromising the objectivity of their intervention and the accuracy and fairness of the rating system.

Besides, the use of human assistance is a rather resource-consuming task and we noticed that the instructors seemed to influence the outcome to a high degree. It eventually became a psychometric concern in our case.

Validity and usability in exams

One key issue, related to face validity, is that the performance of standardized patients (human actors) might not be regular, but varies due to human factors (Adamo 2005). Actors or external examiners are trained to react and behave the same way and in the same manner at each session, but the same actors/evaluators cannot perform identically to assess a full course of hundreds of students, resulting in the need for several individual assessors/actors. This opens up the potential for non-standardized assessment, which could interfere with the results more than lack of familiarity with technology.

In contrast, a virtual patient offers a measurement tool that guarantees the regularity and reproducibility of patient behaviour as well as the judgement of the student’s interaction (including provoked reactions and conveyed emotions) with the case over time (given that no assistants are used).

Therefore, we suggest that students should be judged on the basis of predefined scoring rubrics with well defined cut-off points, for ease of administration and grading. For example, a possible grading scheme could be defined in terms of pounded score based on the percentage of relevant questions asked in the history taking, physical examinations and lab tests ordered, as well as correct diagnosis in relation to the number of submitted diagnoses.

**Conclusion**

The very fact that all 110 students volunteered both to use ISP and completed the whole session, indicates that in general they were positive about the use of an ISP-like simulation system for assessment. Findings from the statistical analysis showed that significant differences existed among the students in the perceived usefulness of human assistance and their degree of affective response towards the system. One randomized study (Smith et al. 1995) and one descriptive study (Holm 1996) have shown that females score better than males after a training course in communication skills. Interestingly, these effects could be significantly measured by ISP (means of relevant history questions). Confirming thereby previous observations done by Vanden Brink-Muinen et al. (1998).

The present study indicates that computer-based simulations like ISP are able to present and simulate realistic patient encounters to an acceptable level of complexity and allow differentiation of one student’s performance from another, including gender-related differences. A strength of ISP-like systems is that these virtual patient cases can also be programmed to score automatically and immediately present results of the examination thus saving expensive labour and facility resources.

Therefore, if students are trained beforehand, limiting the need for assistants, we believe that VP cases can be a useful complementary tool for assessing some of the many components of clinical competence, including clinical reasoning skills, but they should always be combined with other methods (Pugh & Youngblood 2002).

**Notes on contributors**

OLIVIER COURTELLE PhD candidate at Karolinska Institutet, senior developer and project manager in Educational Technologies at the VP-Lab, Karolinska Institutet.

ROLF BERGIN (deceased 2006) Formerly Senior Researcher in Medical Education and Simulation at Karolinska Institutet.

DAG STOCKELD Teacher at Karolinska Institutet since 1986. Senior Consultant in Surgery at Dundee University Hospital in Stockholm.

SARI PONZER Professor in Orthopaedics at Karolinska Institutet. Responsible for organizing the OSCE together with other senior lecturers at Karolinska Institutet. Member of the research team, participated in designing the study and commented on the draft paper.

Downloaded By: [Learning Resource Center/Wrs] At: 14:44 29 June 2008
UNO FORS Researcher and teacher at Karolinska Institutet since 1980. Professor in Medical Educational Simulation and Chairman of LIME.

Acknowledgements

The authors wish to thank Dr Staffan Sahlin for adapting and testing the remodelled virtual patient case and allowing us to test the ISP system in the OSCE exam. The authors are also grateful to the 110 students who volunteered to participate in this study. Thanks also to Jacob Bergström for statistical assistance and analysis.

The study was supported by grants from the Wallenberg Global Learning Network (WGLN) and from Karolinska Institutet.

References


More information on ISP: http://ispvl.learninglab.ki.se