International Collection of Virtual Patients - Digitized Education in Europe beyond the pandemic

iCoViP

2020-1-DE01-KA226-HE-005754

IO4 - Results and conclusions from the pilot implementations

Intellectual Output	IO4
Delivery date	November 2022
Status	final
License	BY-NC-ND
Authors	UNIZAR, All partners
Reviewed by	All partners

Co-funded by the Erasmus+ Programme of the European Union

The creation of these resources has been (partially) funded by the ERASMUS+ grant program of the European Union under grant no. 2020-1-DE01-KA226-005754.

Neither the European Commission nor the project's national funding agency DAAD are responsible for the content or liable for any losses or damage resulting from the use of these resources.

1. Introduction

In this report we describe the implementation of the piloting of iCoViP virtual patients (VPs) at partner and external institutions. The planning and setup of the pilots with the CASUS learning system started in February 2022 and we managed to obtain first data during the summer term 2022. Overall, we piloted 85 VPs with 691 learners from nine countries.
The results of these pilots provided important information for the development of our integration guideline (IO3) and also gave partners the possibility to experiment with VPs in different educational settings.

2. Methods

Each partner was responsible for the implementation of the pilot at their institution and selected the number of VPs and topics based on their needs. Partners were supported by UAU and Instruct in terms of technical aspects and the overall implementations were coordinated by the UNIZAR team.
We based the analysis of the pilots on two pillars: (1) we applied three different questionnaires to evaluate the VPs and their integration and (2) we analyzed the session data of learners based on the data stored in CASUS. However, as we wanted to test the VPs in as many different educational and contextual settings as possible, we do not think it is scientifically sound to compare the results from the partners with each other. But, we can draw general conclusions on what worked well and what could be improved at partner institutions or in general.
We uploaded all anonymized data as an Excel file on our website. An identification of actual users or sessions is not possible with this dataset.

2.1 Design and implementation of the questionnaires

We developed three different questionnaires to evaluate (1) individual VPs, (2) the collection of VPs provided to students, and (3) the integration of these VPs within the pilot setting.
At the beginning of this IO, we dedicated several online meetings to the selection of questions for these questionnaires and finally decided to make use of previously published evaluations tools by Huwendiek et al. [1, 2]. The questionnaires include questions about personal data, such as age, gender, and medical school, questions about the VP / VP collection / integration on a 5-item Likert scale, and open questions for comments. Partners then translated the questionnaires they planned to use into their language. These versions were then implemented in LimeSurvey, an online evaluation tool that is hosted by our associate partner Instruct, who provided the technical support for implementing the questionnaires. The final versions of questionnaires can be downloaded from our website (evaluation tools). Partners selected one of the questionnaires to be presented to the students during and after having worked on the VPs. The survey responses were anonymous, there was no transfer of any personal data from CASUS to LimeSurvey.
After a pilot had been completed, UAU exported the survey data from LimeSurvey into Excel and partners analyzed their data quantitatively guided by UNIZAR and JU. Some partners, such as JU also included a qualitative analysis of the free text responses (see 2.3.1).

2.2 Analysis of session data

The CASUS system provides a wide range of data, which is stored in the CASUS database during a VP session. This includes timestamps for each action, responses to questions and scores, and all activities in the student's concept map for clinical reasoning. We discussed with partners which data to select and analyze prior to the pilots and decided to focus on the clinical reasoning, and analyze the following quantitative variables:

- Diagnostic accuracy (i.e., number of correct final diagnosis on first attempt divided by the overall number of maps)
- Final diagnosis from system: Number of maps in which the learner gave up on finding the correct final diagnosis (after at least one attempt) and the solution was provided by the system divided by the overall number of maps.
- Mean confidence (in %) provided by the learners upon submitting their final diagnosis
- Mean number of problems / findings entered by the learners
- Mean number of differential diagnoses entered by the learners
- Mean number of tests entered by the learners
- Mean number of treatment options entered by the learners
- Mean number of connections drawn by the learners
- Summary statement created by learners (yes / no)
- Mean time on task calculated by the system (this is just an approximation especially if users do not logout or are inactive on a card for some time)
- Mean overall score for questions included in the VP

After a pilot had ended, we exported the aforementioned (anonymized) data from the CASUS database and imported it into MS Excel. Incomplete sessions, i.e. without a final diagnosis submitted by the student were removed from the dataset. For the remaining data, we calculated the quantitative variables for each accessed VP in Excel and summarized them for this report (tables 19 - 27). For pilots where many different VPs were used by learners (e.g. the KUM / UAU pilot), we did not calculate the means on a VP-basis, but over all sessions.

2.3 Description of pilot implementations

All pilots were conducted with the CASUS learning system allowing us to create the VPs, but also assemble these in courses and give students access to such a subset of VPs. The following sections provide a tabular overview about the settings of the pilots implemented by the partners.

2.3.1 Pilots at the Jagiellonian University in Krakow

Integration framework	**Description of pilot**
Phase	Students of medicine in 2nd year (pre-clinical)
Allocated Time	Available from March to May 2022. Students decided when to solve before the deadline.
Setting	Asynchronous and self-directed
Relation to other learning activities	Additional learning activity to promote clinical reasoning and proof of activity between the classes
Groups	Individual work
Orientation	Students were introduced to the concept of virtual patients and how to use CASUS in a class of the Telemedicine course the same year. In addition written instructions were sent via email.
Technical integration	Integrated into the Moodle platform of the medical school via the LTI interface.
Virtual patients (VPs)	● Caroline Bach ● Kurt Baier ● Jelena Jakovic ● Emma Kruger ● Karim Murasic ● Marlene Reister ● Leslie Smith
Survey	Questionnaire on the virtual patient collection (Polish version)

Table 1: Description of pilot setting implemented by JU

We distributed the invitation to fill in the VP collection questionnaire via a mailing list with a few reminders. The collection of responses was closed three weeks after the initial invitation. In addition to the quantitative analysis, we carried out a data-driven thematic analysis of the three open-ended answers in the native language of the responders (Polish) and independently added an English translation of students' comments. The coding frame was refined iteratively until coding of the responses left no items in the residual category. Multiple codes could be assigned to each segment. Each code was extended with a modifier indicating the valence of the comment: positive, negative, neutral.

2.3.2 Pilots in Germany implemented by the Universities of Munich and Augsburg

Integration framework	**Description of pilot**
Phase	All medical students in clinical years across Germany
Allocated Time	open from end of April until end of August (summer term 2022), no time restrictions
Setting	Asynchronous and self-directed
Relation to other learning activities	Additional learning activity promoted to students to train their clinical reasoning based on common key symptoms
Groups	Students could work individually or in pairs / groups, there was no specific instruction
Orientation	No dedicated orientation, but we recommended to watch the introductory video
Technical integration	Freely available course accessible via Shibboleth or self-registration. At UAU and KUM: additionally integrated into the medical schools Moodle platforms via the LTI interface.
Virtual patients (VPs)	We provided five courses covering different key symptoms: ● Abdominal pain / Fatigue (11 VPs) ● Chest pain / Dyspnea (12 VPs) ● Vomiting / Diarrhea (10 VPs) ● Fever / Cough (13 VPs) ● Syncope / Headache (12 VPs)
Survey	Questionnaire on the virtual patient collection without personal data (German version)

Table 2: Description of pilot setting implemented by KUM and UAU

2.3.3 Pilots at the Paris Saclay University

Integration framework	**Description of pilot**
Phase	Medical students in year 4 and 5
Allocated Time	One session of 3 hours dedicated to clinical reasoning for each group in May 2022
Setting	Synchronous, in presence of teachers and with immediate support from teachers.
Relation to other learning activities	Additional learning activity promoted to students to train their clinical reasoning based on past courses validated
Groups	Students could work individually or in pairs, there was no specific instruction
Orientation	15 minutes introduction to clinical reasoning and the virtual patients by teachers and a step by step introduction of one VP. No preliminary work required.
Technical integration	Access on the university computers via Shibboleth to ups.casus.net
Virtual patients (VPs)	We selected six VPs in French and verified by local teachers on topics studied in the last 2 years by the students. The selected VPs were: ● Maria Bauer ● Robert Wallner ● Caroline Bach ● Britta Ohland ● Erik Marte ● Krystyna Kowalczyk
Survey	Questionnaire on the virtual patient collection (French version)

Table 3: Description of pilot setting implemented by UPS

2.3.4 Pilots at the University of Zaragoza

Integration framework	**Description of pilot**
Phase	Medical students in their clinical years (from 4th to 6th year) who voluntarily agreed to participate in the study.
Allocated Time	Last week of April 2022 (25th April at the Faculty) but more time was given for those who wanted to do the activity remote from home.
Setting	Synchronous, in presence of teachers and with immediate support from teachers. All participants who completed the VPs (minimum of four) and answered two questionnaires were given a voucher for the cafeteria of the Faculty of Medicine.
Relation to other learning activities	Additional learning activity promoted to medical students to train their clinical reasoning based on common key symptoms.
Groups	Students could work individually or in pairs, there was no specific instruction.
Orientation	15 minutes introduction to clinical reasoning and the VPs from teachers and step by step following one VP as an example. No preliminary work required, but watching the explanatory video was recommended.
Technical integration)	Students used their own computers and access via self-registration.
Virtual patients (VPs)	The following nine virtual patients were included: ● Maria Bauer ● Alexandra May ● Yara Mahmoudi ● Miko Sasagawa ● Robert Wallner ● Karim Murasic ● Bernhard Hinkel ● Anja Winterberg ● Ulrike Birnbaum
Survey	Questionnaire on the virtual patient collection (Spanish version) and individual virtual patients. Once the virtual patients had been completed, students had to evaluate the one case he/she found most interesting.

Table 4: Description of pilot setting implemented by UNIZAR

2.3.5 International Pilots implemented by the University Hospital of Munich

In addition to the pilots with German students, KUM implemented three additional pilots with students from the Asia-Pacific region, Latin America, and Lund University. Our associate partners EUGLOH and CIH supported us in initializing these pilots. The settings are described in the following tables (5-7)

Integration framework	**Description of pilot 1**
Phase	Medical students from the Asia-Pacific region. Students from the University of Nagoya (Japan), Korea University, Chinese University of Hong Kong, Monash University (Australia), University of Bologna (Italy) and the LMU Munich currently enrolled in the Joint Online Clinical Case Discussions activity (JOCCD 2022) were invited to participate in the pilot.
Allocated Time	Available from May 22nd until July 5th
Setting	Asynchronous and self-directed; Students were asked to complete a minimum of 5 VPs by July 5th 2022 to receive a certificate of completion.
Relation to other learning activities	Additional learning activity promoted to students to train their clinical reasoning based on common key symptoms.
Groups	Students could work individually or in pairs / groups, there was no specific instruction
Orientation	No dedicated orientation, but we recommended to watch the introductory video
Technical integration	Access was via self-registration
Virtual patients (VPs)	We provided access to 23 virtual patients covering the key symptoms of abdominal pain, fatigue, chest pain, and dyspnea
Survey	Questionnaire on the virtual patient collection without personal data (English version)

Table 5: Description of pilot setting implemented by KUM for students from the Asian-Pacific region

Integration framework	**Description of pilot 2**
Phase	Medical students at the University of Lund (Sweden)
Allocated Time	Available from October 1st to October 31st 2022.
Setting	Asynchronous and self-directed, students received a certificate for completing a minimum of 5 virtual patients of their choice from within the offered collection of VPs.
Relation to other learning activities	Additional learning activity promoted to students to train their clinical reasoning as an addition to their regular rotations in surgery and infectious diseases.
Groups	Students could work individually or in pairs / groups, there was no specific instruction
Orientation	No dedicated orientation, but we recommended to watch the introductory video
Technical integration	Access was via self-registration
Virtual patients (VPs)	We provided access to 30 virtual patients for infectious diseases and 17 virtual patients for surgery.
Survey	Questionnaire on the virtual patient collection without personal data (English version)

Table 6: Description of pilot setting implemented by KUM for students from Lund University

Integration framework	**Description of pilot 3**
Phase	Medical students from the Latin American Region
Allocated Time	Available from October 10th to October 31st 2022.
Setting	Asynchronous and self-directed, students received a certificate for completing five virtual patients.
Relation to other learning activities	Additional learning activity promoted to students to train their clinical reasoning based on common key symptoms.
Groups	Students could work individually or in pairs / groups, there was no specific instruction
Orientation	No dedicated orientation, but we recommended to watch the introductory video
Technical integration	Access was via self-registration
Virtual patients (VPs)	We provided access to 28 virtual patients covering the key symptoms of abdominal pain, fatigue, chest pain, and dyspnea
Survey	Questionnaire on the virtual patient collection (Spanish and Portuguese version)

Table 7: Description of pilot setting implemented by KUM for students from Latin America

2.3.6 Pilots at the University of Porto

Integration framework	**Description of pilot**
Phase	Medical students in their clinical years (from 4th to 6th year) who voluntarily agreed to participate in the study.
Allocated Time	Two first weeks of October 2022 with more time was given for those who wanted to do the activity remote from home.
Setting	Asynchronous and self-directed, students received a certificate for completing five virtual patients.
Relation to other learning activities	Additional learning activity promoted to medical students to train their clinical reasoning based on common key symptoms.
Groups	Students could work individually or in pairs, there was no specific instruction.
Orientation	15 minutes introduction to clinical reasoning and the VPs from teachers and step by step following one VP as an example. No preliminary work required, but watching the explanatory video is recommended.
Technical integration	Students used their own computers and access via self-registration.
Virtual patients (VPs)	The following virtual nine patients were provided: ● Maria Bauer ● Alexandra May ● Yara Mahmoudi ● Miko Sasagawa ● Robert Wallner ● Karim Murasic ● Bernhard Hinkel ● Anja Winterberg ● Ulrike Birnbaum
Survey	Questionnaire on the virtual patient collection (Portuguese version).

Table 8: Description of pilot setting implemented by UPORTO

2.3.7 Pilots in Ukraine implemented by Augsburg

Integration framework	**Description of pilot**
Phase	Medical students at the Bukovinian State Medical University (Chernivtsi, Ukraine)
Allocated Time	Available from Oct 4th to Oct 31st, 2022 (data collection continued until end of Dec 2022)
Setting	Synchronous learning activity in form of seminars / case studies led by a teacher
Relation to other learning activities	Learning activity offered to students to train their clinical reasoning as an addition to their regular rotations in the hospital.
Groups	Students work as a group to solve the cases
Orientation	Teachers presented the concept mapping approach to visualize clinical reasoning
Technical integration	Virtual patients were presented on a digital whiteboard during seminars
Virtual patients (VPs)	We provided access to 23 virtual patients for infectious, vascular, immunologic, metabolic, congenital, traumatic, and neoplastic diseases
Survey	Questionnaire on the virtual patient collection (Ukrainian version)

Table 9: Description of pilot setting implemented by UAU for Ukrainian students

3. Results

3.1 Overview of pilot results

The following table provides an overview of some fundamental data, such as the number of participants, sessions, and survey responses for each pilot implementation (table 10).

	No Participants (>=1 completed session)	Total sessions	Completed Sessions	Completed surveys	Response
JU	238	1470	1450	75	31%
Germany	55	556	364	-	-
UPS	57 (47)	218	183	47	87%
UNIZAR	65	338	318	29 (collection) + 67 (indiv. VPs)	45%
UPORTO	48 (36)	100	69	21	44%
Asia Pacific	5	21	17	-	-
Ukraine	132 (12*)	22	20	75	57%
Lund University	28 (20)	87	69	7	25%
Latin America	63 (45)	277	257	-	-
Total	**691**	**3.089**	**2.727**	**321**	

*Table 10: Overview of pilot results; (*VPs mostly used in a seminar-style setting with VP shown on digital whiteboard (therefore not many sessions)).*

3.2 Details of survey results

In the following sections, we summarize the analysis results of survey responses from each pilot implementation.

3.2.1 Results of the Jagiellonian University in Kraków

Overall, the vast majority of students were satisfied with the learning experience of using iCoViP virtual patients (85% agreed or strongly agreed). In particular, the students were satisfied with the active engagement in updating the thinking process while new information became available (88%) and building their differential diagnosis. Slightly less positive, but still on a very good level was the evaluation of the authenticity of the cases (82%). After the classes, 81% of students felt confident to explain what clinical reasoning is. Relatively most problematic was the functionality provided by the concept map. Yet still, the majority of students agreed that the concept maps helped them in structuring their thoughts (67%), were easy to use (65%) and provided good feedback (57%).

	VP collection (N= 75 completed)	Strongly agree	Somewhat agree	Neutral	Somewhat disagree	Strongly disagree		
Id	Question	5	4	3	2	1	mean	SD
1	While working on the virtual patient collection, I felt I had to make the same decisions a doctor would make in real life.	14.7%	66.7%	5.3%	9.3%	4.0%	3.79	0.95
2	The concept map helped structuring my thoughts.	21.3%	45.3%	9.3%	13.3%	10.7%	3.53	1.27
3	While working through the virtual patient collection, I was actively engaged in thinking about which findings supported or refuted each diagnosis in my differential diagnosis.	36.0%	50.7%	5.3%	5.3%	2.7%	4.12	0.93
4	While working through the virtual patient collection, I was actively engaged in updating my thinking process as new information became available.	46.7%	41.3%	5.3%	4.0%	1.3%	4.30	0.86
5	The feedback I received in the concept map was helpful in enhancing my clinical reasoning in this virtual patient collection.	21.3%	36.0%	22.7%	16.0%	4.0%	3.55	1.12
6	I thought the concept map was easy to use.	21.3%	44.0%	13.3%	16.0%	5.3%	3.60	1.15
7	I think I am able to explain to my colleagues what clinical reasoning is.	32.0%	49.3%	10.7%	5.3%	2.7%	4.03	0.94
8	Overall, working through the virtual patient collection was a worthwhile learning experience.	49.3%	36.0%	12.0%	1.3%	1.3%	4.31	0.84

Table 11: Quantitative survey results

In general, our qualitative analysis confirmed the findings from the quantitative description but delivered more depth in understanding students' opinions:

		Valence	
		+	**-**
Content	1.1 Authenticity	5	1

	1.2 Case selection	8	4
	1.3 Medical data	3	2
Learning	2.1 Curricular integration	4	5
	2.2 Feedback	6	6
	2.3 Interactivity	12	5
	2.4 Learning style	3	2
	2.5 Outcomes	10	0
	2.6 Safe environment	8	0
	2.7 Self-assessment	11	3
	2.8 Guidance	3	4
Tech	3.1 Controlled vocabulary	0	9
	3.2 Usability	2	7
	3.3 Concept map	1	4
Other	4.1 General	6	0
	4.2 Unclear	1	3
	4.3 Empty	4	0

Table 12: Summary of qualitative survey results

3.2.2 Results of pilots in Germany implemented by the Universities of Munich and Augsburg

Despite having attached the course-level questionnaire to the course, we did not receive any responses to the survey. This is not uncommon in such settings as students do not see the immediate benefit of evaluation like in more integrated settings. Therefore, we will focus for this pilot setting on analyzing the learner sessions across all VPs (see 3.3.2).

3.2.3 Results of the Paris Saclay University

Overall, the majority of students were happy with the learning experience. They actively and happily engaged in this new way to learn clinical reasoning. They debriefed their experience with the teachers at the end of the session and were satisfied with the opportunity to learn clinical reasoning with a different platform than paper and pen, but also with the tools provided here. An overview about the quantitative survey results is shown in table 13.
The absence of an official scoring was seen as an advantage. They reported that evaluating their skills with this tool would be difficult as the requirement they felt in terms of clinical knowledge and habits of providing medicine was different to the one they were used to and

the one taught at the university. This is one disadvantage of the multicultural project. The use of the clinical reasoning tool was difficult to grasp at first but as they were getting used to it, the technical issues decreased.

	VP collection (N= 47 completed)	Strongly agree	Somewhat agree	Neutral	Somewhat disagree	Strongly disagree		
Id	Question	5	4	3	2	1	Mean	SD
1	While working on the virtual patient collection, I felt I had to make the same decisions a doctor would make in real life.	10.9%	67.3%	10.9%	10.9%	0	3.78	0.74
2	The concept map helped structuring my thoughts	6.5%	52.2%	26.1%	10.9%	4.3%	3.46	0.96
3	While working through the virtual patient collection, I was actively engaged in thinking about which findings supported or refuted each diagnosis in my differential diagnosis.	37.0%	50.0%	10.8%	2.2%	0	4.22	0.75
4	While working though the virtual patient collection, I was actively engaged in updating my thinking process as new information became available.	41.3%	45.7%	10.8%	2.2%	0	4.26	0.77
5	The feedback I received in the concept map was helpful in enhancing my clinical reasoning in this virtual patient collection	6.8%	40.9%	31.8%	18.2%	2.3%	3.32	1.02
6	I thought the concept map was easy to use	13%	26.2%	21.7%	30.4%	8.7%	3.04	1.37
7	I think I am able to explain to my colleagues what clinical reasoning is	28.3%	63%	8.7%	0	0	4.5	0.55
8	Overall, working through this virtual patient collection was a worthwhile learning experience.	52.2%	45.6%	2.2%	0	0	4.5	0.55

Table 13: Quantitative results from the VP collection survey (n=47)

3.2.4 Results of the University of Zaragoza

We are presenting two tables from the two questionnaires students filled in the pilot we had at the end of April in Zaragoza. Table 14 summarized the results from the individual VP used in our course with 67 responses and table 15 the results from the individual VP questionnaire with 28 responses. Despite the low response, in our opinion, we were able to collect the general opinion of the participating students.

Overall, most of the students were satisfied with the learning experience of using iCoViP virtual patients (76.7% agreed or strongly agreed). In particular, the students were satisfied with the active engagement in updating the thinking process while new information became available (96.5%) and building their differential diagnosis. Slightly less positive was the evaluation of the authenticity of the cases (67.9%). After the classes, 89.3% of students felt confident to explain what clinical reasoning is. The most problematic aspect was the functionality provided by the concept map. Not even half of the students agreed the concept maps helped them in structuring their thoughts (46.3%), or considered concepts maps were easy to use (39.3%) and provided good feedback (50.0%). In fact, between one third and one half of the students were dissatisfied with the functionality of the concept maps.

	Individual VP (N=67 completed)	Strongly agree	Somewhat agree	Neutral	Somewhat disagree	Strongly disagree		
Id	Question	5	4	3	2	1	Mean	SD
1	While working on this virtual patient, I felt I had to make the same decisions a doctor would make in real life.	10.4%	62.7%	10.4%	10.4%	5.9%	3.61	1.01
2	While working on this virtual patient, I felt I were the doctor caring for this patient.	9.0%	59.7%	16.4%	7.5%	7.5%	3.55	1.09
3	While working through this virtual patient, I was actively engaged in gathering the information (e.g., history questions, physical exams, lab tests) I needed, to characterize the patient's problem.	17.9%	50.7%	16.4%	10.4%	4.5%	3.67	1.04
4	While working through this virtual patient, I was actively engaged in revising my initial image of the patient's problem as new information became available.	25.4%	59.7%	6.0%	7.5%	1.5%	4.05	0.79
5	While working through this virtual patient, I was actively engaged in creating a short summary of the patient's problem using medical terms.	7.5%	56.7%	17.9%	16.4%	1.5%	3.52	0.91

6	While working through this virtual patient, I was actively engaged in thinking about which findings supported or refuted each diagnosis in my differential diagnosis.	25.4%	61.2%	9.0%	3.0%	1.5%	4.06	0.77
7	I felt that the virtual patient was at the appropriate level of difficulty for my level of training.	19.4%	52.2%	10.4%	14.9%	3.0%	3.70	1.04
8	The questions I was asked while working through this virtual patient were helpful in enhancing my clinical reasoning in this virtual patient.	17.9%	55.2%	14.9%	9.0%	3.0%	3.76	0.95
9	The feedback I received was helpful in enhancing my diagnostic reasoning in this virtual patient.	20.9%	52.2%	13.4%	7.5%	6.0%	3.75	1.06
10	After completing this virtual patient, I feel better prepared to confirm a diagnosis and exclude differential diagnoses in a real life patient with this complaint.	13.4%	58.2%	20.4	6.0%	1.5%	3.76	0.82
11	After completing this virtual patient I feel better prepared to care for a real life patient with this complaint.	1.5%	61.2%	13.4%	14.9%	1.5%	3.61	0.90
12	Overall, working through this virtual patient was a worthwhile learning experience.	22.4%	53.7%	10.4%	10.4%	3.0%	3.82	0.99

Table 14: Quantitative results from the individual VP questionnaire (n=67).

	VP collection (N=28 completed)	Strongly agree	Somewhat agree	Neutral	Somewhat disagree	Strongly disagree		
Id	Question	5	4	3	2	1	Mean	SD
1	While working on this virtual patient, I felt I had to make the same decisions a doctor would make in real life.	17.9%	50.0%	14.3%	10.7%	7.1%	3.60	1.13
2	The concept map helped structuring my thoughts	14.2%	32.1%	10.7%	14.3%	28.6%	2.89	1.49
3	While working through this virtual patient, I was actively engaged in thinking about which findings supported or refuted each diagnosis in my differential diagnosis.	7.1%	85.7%	3.6%	3.6%	0	3.96	0.50
4	While working though the virtual patient collection, I was actively engaged in updating my thinking process as new information became available.	17.9%	78.6%	0	3.6%	0	4.10	0.56
5	The feedback I received in the concept map was helpful in enhancing my clinical reasoning in this virtual patient collection	17.9%	32.1%	7.1%	28.6%	14.3%	3.10	1.39
6	I thought the concept map was easy to use	10.7%	28.6%	17.9%	21.4%	21.4%	2.85	1.35
7	I think I am able to explain to my colleagues what clinical reasoning is	21.4%	67.9%	7.1%	3.6%	0	4.07	0.66
8	Overall, working through this virtual patient collection was a worthwhile learning experience.	10.7%	60.7%	14.3%	14.3%	0	3.67	0.86

Table 15: Quantitative results from the VP collection questionnaire (n=28).

3.2.5 Results of the international pilots implemented by the University Hospital of Munich

Although the pilots implemented by the University of Munich spanned a total of 11 institutions, only 7 students from Lund University completed the final evaluation questionnaire (Table 16). On average, students at Lund University showed high agreement items related to real-life decision-making (mean of 4.29 where 5 corresponds to strongly agree), engagement in thinking about differential diagnoses (mean 4.14) and self-confidence in explaining clinical reasoning to peers (mean 4.43). Conversely, the scores for items related to the concept map were lower in terms of usefulness for structuring thoughts (mean 3.43), feedback received in the concept map (mean 2.00) and whether the concept map was easy to use (mean 3.86).

	VP collection (N= 7 completed)	Strongly agree	Somewhat agree	Neutral	Somewhat disagree	Strongly disagree	
Id	Question	5	4	3	2	1	Mean
1	While working on this virtual patient, I felt I had to make the same decisions a doctor would make in real life.	29%	71%	0	0	0	4.29
2	The concept map helped structuring my thoughts	0	57%	29%	14%	0	3.43
3	While working through this virtual patient, I was actively engaged in thinking about which findings supported or refuted each diagnosis in my differential diagnosis.	57%	14%	14%	14%	0	4.14
4	While working though the virtual patient collection, I was actively engaged in updating my thinking process as new information became available.	71%	29%	0	0	0	4.71
5	The feedback I received in the concept map was helpful in enhancing my clinical reasoning in this virtual patient collection	14%	29%	0	0	57%	2.43
6	I thought the concept map was easy to use	0	14%	29%	0	57%	2.00
7	I think I am able to explain to my colleagues what clinical reasoning is	14%	71%	0	14%	0	3.86
8	Overall, working through this virtual patient collection was a worthwhile learning experience.	43%	57%	0	0	0	4.43

Table 16: Quantitative results from the VP collection questionnaire (n=7) with students from Lund University.

3.2.6 Results of the University of Porto

A total of 21 students completed the virtual patient collection questionnaire (Table 17). Overall, most students agreed or strongly agreed with the statements of the questionnaire. In fact, such agreement was higher than two-thirds for all questions except one ("I thought the concept map was easy to use"). The statement with the highest agreement (100%) was "While working though the virtual patient collection, I was actively engaged in updating my thinking process as new information became available". However, the precision of the obtained estimates is low due to the small sample size.

	VP collection (N= 21 completed)	Strongly agree	Somewhat agree	Neutral	Somewhat disagree	Strongly disagree		
Id	Question	5	4	3	2	1	Mean	SD
1	While working on this virtual patient, I felt I had to make the same decisions a doctor would make in real life.	38.1%	47.6%	9.5%	0%	4.8%	4.14	0.96
2	The concept map helped structuring my thoughts	57.1%	33.3%	9.5%	0%	0%	4.48	0.68
3	While working through this virtual patient, I was actively engaged in thinking about which findings supported or refuted each diagnosis in my differential diagnosis.	61.9%	28.6%	4.8%	4.8%	0%	4.48	0.81
4	While working though the virtual patient collection, I was actively engaged in updating my thinking process as new information became available.	66.7%	33.3%	0%	0%	0%	4.67	0.48
5	The feedback I received in the concept map was helpful in enhancing my clinical reasoning in this virtual patient collection	61.9%	19.0%	4.8%	9.5%	4.8%	4.24	1.22
6	I thought the concept map was easy to use	47.6%	14.3%	14.3%	19.0%	4.8%	3.81	1.36
7	I think I am able to explain to my colleagues what clinical reasoning is	33.3%	38.1%	19.0%	4.8%	0%	4.05	0.89

8	Overall, working through the virtual patient collection was a worthwhile learning experience.	57.1%	28.6%	9.5%	0%	0%	4.50	0.69

Table 17: Quantitative results from the VP collection questionnaire (n=21)

3.2.6 Results of pilots in Ukraine implemented by Augsburg

	VP collection (N= 75 completed)	Strongly agree	Somewhat agree	Neutral	Somewhat disagree	Strongly disagree	
Id	Question	5	4	3	2	1	Mean
1	While working on this virtual patient, I felt I had to make the same decisions a doctor would make in real life.	71.6%	25.7%	2.7%	0	0	4.68
2	The concept map helped structuring my thoughts	73.0%	20.3%	2.7%	4.1%	0	4.61
3	While working through this virtual patient, I was actively engaged in thinking about which findings supported or refuted each diagnosis in my differential diagnosis.	74.3%	20.3%	4.1%	1.4%	0	4.68
4	While working though the virtual patient collection, I was actively engaged in updating my thinking process as new information became available.	87.8%	9.5%	0	2.7%	0	4.82
5	The feedback I received in the concept map was helpful in enhancing my clinical reasoning in this virtual patient collection	78.4%	17.6%	2.7%	1.4%	0	4.73
6	I thought the concept map was easy to use	67.6%	25.7%	2.7%	1.4%	0	4.57
7	I think I am able to explain to my colleagues what clinical reasoning is	72%	26.7%	1.3%	0	0	4.71
8	Overall, working through this virtual patient collection was a worthwhile learning experience.	90.7%	9.3%	0	0	0	4.91

Table 18: Quantitative results from the VP collection questionnaire (n=75).

3.3 Analysis of session data

For the analysis of session data we included only completed sessions, i.e sessions in which students submitted a final diagnosis.

3.3.1 Results of the Jagiellonian University in Kraków

In this pilot course, 238 participants completed between one and seven VPs (mean=6.7).

	CB	**KM**	**EK**	**LS**	**KB**	**MR**	**JJ**
Completed maps	97	208	233	206	233	233	233
Diagnostic accuracy (%)	83.5	77.9	80.7	0	82.8	94.4	85.8
Final diagnosis from system (%)	8.3	14.9	9.4	100	13.7	3.9	9.0
mean confidence (%)	53.5	66.6	65.9	66.8	64.6	70.3	62.1
mean No problems / findings	9.2	5.1	6.6	8.5	6.0	6.1	5.3
mean No differentials	3.6	4.2	3.9	4.8	4.0	3.1	3.6
mean No tests	4.5	4.7	5.0	4.5	4.6	3.6	3.2
mean No treatments	1.6	1.1	2.0	1.5	1.3	1.5	1.6
mean No connections	6.3	5.5	5.6	5.0	4.3	4.2	3.3
Summary statement created (%)	69.1	63.0	60.1	56.8	60.5	60.5	62.6
mean time on task (min)	29.9	25.6	24.5	26.1	18.8	15.6	16.0
mean score on questions (%)	59.2	46.7	53.0	59.8	57.2	37.4	30.9

Table 19: Summary of VP session for the VPs: CB=Caroline Bach, KM=Karim Murasic, EK=Emma Kruger, LS=Leslie Smith, KB=Kurt Baier, MR=Marlene Reister, JJ=Jelena Jakovic

3.3.2 Results of pilots in Germany implemented by the Universities of Munich and Augsburg

In this pilot course 76 different VPs have been completed (minimum 2 sessions / VP, maximum 19 sessions / VP). As the numbers are too low for a detailed analysis, we analyzed this course over all VPs and included an analysis of VPs with more than 10 sessions. The 55 learners completed on average 6.6 VPs (minimum=1, maximum=65).

	All VPs	**VN**	**CZ**	**HP**	**RB**	**LT**	**KG**
Completed maps	364	19	12	15	11	14	11
Diagnostic accuracy (%)	45.6	21.1	75.0	26.7	81.8	78.6	72.7
Final diagnosis from system (%)	33.0	57.9	16.7	66.7	18.2	14.3	72.2
mean confidence (%)	43.1	52.4	69.1	47.2	60.0	48.7	46.5
mean No problems / findings	4.7	4.0	4.0	6.2	7.6	4.9	1.9
mean No differentials	3.8	3.6	3.7	7.4	3.1	4.0	1.9

mean No tests	3.6	1.7	4.3	5.3	4.1	3.6	2.1
mean No treatments	1.1	0.6	3.0	0.4	2.3	1.0	1.2
mean No connections	0.2	0.3	0.6	0.6	0.7	0.2	0
Summary statement created (%)	37.6	31.6	33.3	46.7	45.5	57.1	36.4
mean time on task (min)	16.6	12.8	11.9	24.1	16.5	12.5	12.6

Table 20: Summary of sessions for all 74 VPs in the pilot course and VPs with more than 10 sessions: VN=Valentina Nadler, CZ=Carl Zimmermann, HP=Hedwig Paul, RB=Robert Baley, LT=Leoni Teichert, KG=Klara Gebhart

3.3.3. Results of the Paris Saclay University

In this pilot course, 47 participants completed between three and six VPs (mean=4.7). The last VP of the course (Robert Wallner) was only completed by four students, so we did not include it in the analysis.

	BO	**AW**	**CB**	**EM**	**KK**	**MB**
Completed maps	26	44	43	34	8	24
Diagnostic accuracy (%)	88.5	52.3	69.8	61.8	12.5	75.0
Final diagnosis from system (%)	3.9	4.6	14.0	8.8	87.5	4.2
mean confidence (%)	75.3	86.6	75.4	73.8	34.9	78.9
mean No problems / findings	11.2	7.8	10.8	7.4	12.4	9.3
mean No differentials	6.9	6.1	4.9	4.6	4.5	7.8
mean No tests	8.6	8.0	7.8	4.9	7.0	7.3
mean No treatments	3.6	1.4	2.3	1.7	0.3	1.8
mean No connections	11.7	12.8	13.4	8.2	10.5	9.5
Summary statement created (%)	84.6	88.6	83.7	76.5	75.0	87.5
mean time on task (min)	63.0	28.7	25.2	18.1	18.3	54.2
mean score on questions (%)	47.8	50.0	68.6	78.5	30.3	67.9

Table 21: Quantitative session data: VPs in same order as provided in the course - BO=Britta Ohland, AW=Anja Winterberg, CB=Caroline Bach, EM=Erik Marte, KK=Krystyna Kowalczyk, MB=Maria Bauer

3.3.4 Results of the University of Zaragoza

In this pilot course in Zaragoza, 65 participants completed at least one virtual patient (min=1, max=9, mean=4.9). The following table provides an overview of the session data of

participants in the pilot course. The table only includes VP sessions (n=318) in which participants submitted a final diagnosis.

	MB	AM	YM	MS	RW	KM	BH	AW	UB
Completed maps	66	55	53	45	26	22	17	19	15
Diagnostic accuracy (%)	39.4	90.9	34.0	68.9	19.2	59.1	60.9	47.4	13.3
Final diagnosis from system (%)	15.2	1.8	26.4	2.2	46.2	22.7	35.3	21.1	73.3
mean confidence (%)	73.4	78.2	67.3	71.8	54.7	62.4	50.5	73.5	53.3
mean No problems / findings	5.0	4.7	5.5	5.4	3.6	2.1	2.3	3.9	2.7
mean No differentials	6.1	3.4	5.2	3.7	3.4	4.0	2.5	3.7	3.7
Avg. No tests	5.3	2.5	4.6	2.5	2.9	4.2	2.4	4.4	3.3
mean No treatments	2.5	1.4	2.1	1.7	1.0	1.1	2.1	1.5	3.0
mean No connections	3.3	2.5	2.6	3.0	1.7	2.7	1.7	2.0	1.1
Summary statement created (%)	60.6	52.7	49.1	46.7	30.8	50.0	52.9	47.4	53.3
mean time on task (min)	35.2	17.2	16.2	18.1	15.5	26.4	16.3	15.7	14.8
mean score on questions (%)	60.2	87.0	60.2	72.7	43.6	57.6	35.7	55.6	51.4

Table 22: Quantitative session data: VPs in same order as provided in the course - AM=Alexandra May, AW=Anja Winterberg, BH=Bernhard Hinkel, KM=Karim Murasic, MB=Maria Bauer, MS=Miko Sasagawa, RW=Robert Wallner, UB=Ulrike Birnbaum, YM=Yara Mahmoudi

3.3.5 Results of the international pilots implemented by the University Hospital of Munich

3.3.5.1 Pilot with students from the Asia-Pacific region

In this pilot course organized by the University Hospital of Munich, 10 participants from the Asia-Pacific region (mainly Hong Kong and Australia) participated and 5 completed at least one VP. Table 23 shows the summarized results of the pilot with the students from the Asia-Pacific region. From the 23 virtual patients in the course, students accessed and completed 11. Overall, we recorded 17 sessions, so instead of a detailed analysis on a VP basis, we summarize the results across all VPs in the following table.

	All VPs
Completed maps	17
Diagnostic accuracy (%)	52.9
Final diagnosis from system (%)	23.5
Mean confidence (%)	48.4
Mean No problems / findings	6.7
Mean No differentials	8.0
Mean No tests	5.7
Mean No treatments	1.9
Mean No connections	3.7
Summary statement created (%)	76.5
Mean time on task (min)	36.0
Mean score on questions (%)	57.0

Table 23: Quantitative session data for the pilot in the Asia-Pacific region; VPs with at least one completed map: Isabel Schuster, Ahmet Ünal, Alexandra May, Clark Wilman, Hedwig Paul, Isabel Schuster, Josef Schröder, Julia Biederman, Karim Murasic, Leslie Smith, Norman Jacobs, Tim Wagemann.

3.3.5.2 Pilot with students from Latin America

In this pilot course organized by the University Hospital of Munich, 63 participants from Latin America participated and 45 completed at least one VP. Table 24 shows the summarized results of the pilot. From the 28 virtual patients in the course, students accessed and completed 25. Overall, we recorded 257 completed sessions and we report VPs with at least 10 completed maps separately and in addition the overall results for all 25 VPs.

	AM	**EK**	**FR**	**FB**	**IB**	**JJ**	**JMS**	**KR**	**All VPs**
Completed maps	12	14	33	26	21	12	19	13	257
Diagnostic accuracy (%)	91.7	64.3	36.4	92.3	42.9	25.0	15.8	69.2	46.7
Final diagnosis from system (%)	0	14.3	48.5	0	47.6	58.3	63.2	7.7	32.3
Mean confidence (%)	82.5	66.6	63.4	74.3	63.2	59.4	63.9	74.1	66.9
Mean No problems / findings	5.7	5.7	5.7	6.6	3.2	4.7	9.8	4.4	6.0
Mean No differentials	4.0	6.1	3.9	3.1	4.3	6.3	6.4	3.6	5.1
Mean No tests	3.3	4.7	2.8	2.7	3.1	3.5	5.6	3.1	4.0

Mean No treatments	1.9	1.9	1.1	2.0	2.0	3.3	1.2	2.8	1.8
Mean No connections	2.3	2.9	3.2	2.8	4.1	4.1	4.7	3.3	3.0
Summary statement created (%)	100	78.6	78.8	84.6	95.2	83.3	89.5	84.6	89.5
Mean time on task (min)	19.2	30.0	32.3	24.0	26.1	25.8	41.0	21.6	28.3
Mean score on questions (%)	76.7	-	19.4	47.1	-	50.0	50.8	75.0	55.2

Table 24: Overview of results from pilot with students from Latin America. AM: Alexandra May, EK=Elisabeth Kagerl, FR=Francois Robin, FB=Frank Beck, IB=Irene Bach, JJ=Jelena Jakovic, JMS=Jose Maria Saenz, KR=Katharina Rigger.

3.3.5.2 Pilot with students from Lund University, Sweden

In this pilot course organized by the University Hospital of Munich, 28 participants from Lund University participated and 20 completed at least one VP. Table 25 shows the summarized results of the pilot. From the 32 virtual patients in the courses, students accessed and completed 19. Overall, we recorded 69 completed sessions, so instead of a detailed analysis on a VP basis, we summarize the results across all VPs in the following table.

	VPs Infectious	**VPs surgery**
Completed maps	12	57
Diagnostic accuracy (%)	91.7	71.9
Final diagnosis from system (%)	0	17.5
Mean confidence (%)	60.5	63.2
Mean No problems / findings	6.3	4.0
Mean No differentials	6.8	4.7
Mean No tests	4.8	3.0
Mean No treatments	1.5	0.8
Mean No connections	0	0.2
Summary statement created (%)	83.3	47.4
Mean time on task (min)	18.8	16.6
Mean score on questions (%)	59.4	51.9

Table 25: Overview of results from pilot with students from Lund University in Sweden (n=5 infectious, n=15 surgery).

3.2.6 Results of pilots in Porto

In this pilot course organized by the Porto University, 48 students participated and 36 completed at least one VP. Table 26 shows the summarized results of the pilot. All 9 virtual patients in the course have been completed by at least one student. Overall, we recorded 69 completed sessions, so instead of a detailed analysis on a VP basis, we summarize the results across all VPs in the following table.

	All VPs
Completed maps	69
Diagnostic accuracy (%)	59.4
Final diagnosis from system (%)	30.4
Mean confidence (%)	57.7
Mean No problems / findings	5.3
Mean No differentials	3.4
Mean No tests	3.9
Mean No treatments	1.1
Mean No connections	0.6
Summary statement created (%)	26.1
Mean time on task (min)	22.8
Mean score on questions (%)	51.7

Table 26: Quantitative session data; VPs with at least one completed map: Maria Bauer, Alexandra May, Yara Mahmoudi, Miko Sasagawa, Robert Wallner, Karim Murasic, Bernhard Hinkel, Anja Winterberg, Ulrike Birnbaum.

3.2.7 Results of pilots in Ukraine implemented by Augsburg

In this pilot course 132 students participated and jointly in synchronous meetings worked on 11 VPs. Table 27 shows the summarized results of this pilot. Due to the synchronous setting, which focused more on discussing aspects of the case, we only report selected aspects.

	All VPs
Completed maps	20
Diagnostic accuracy (%)	-
Final diagnosis from system (%)	-
Mean confidence (%)	-

Mean No problems / findings	7.0
Mean No differentials	5.3
Mean No tests	7.4
Mean No treatments	2.3
Mean No connections	0.7
Summary statement created (%)	-
Mean time on task (min)	78.9
Mean score on questions (%)	72.2

Table 27: Overview of results from pilot with students from Ukraine. VPs with at least one completed map: Yara Mahmoudi, Miko Sasagawa, Dominik Maller, Miray Günal, Veronika Heidemann, Gerald Fuchs, Tymon Ogryzek, Franciszek Kizior, Ana Bento, Laurenz Schmidt, Moritz Hager.

4. Discussion and Conclusions

Overall, our survey results show a general satisfaction of learners with our virtual patient collection. For example, for all pilots more than 70% of the participants rated the VP collection as a worthwhile learning experience (strongly and somewhat agree). We believe this shows that the iCoViP VP collection is applicable in many different settings as we deliberately designed the pilots to cover a wide variety of settings. We also see differences between settings. For example, students appreciated (strongly and somewhat agree) the concept mapping approach for structuring their thoughts in Porto (90.4%) and Ukraine (93.3%), but less in Zaragoza (46.3%). However, due to our different settings, we can only hypothesize about reasons for such differences, which could be due to differing quality of instruction or prior experience, knowledge, or motivation of students. In general, we see a need at all institutions for accompanying the introduction of virtual patients for clinical reasoning with a basic training in clinical reasoning and approaches to it, such as concept mapping. Such a combination of basic clinical reasoning education and training with virtual patients is for example realized in the DID-ACT project, which we also further describe in our integration guideline.

Looking into the session data (learning analytics), we see, overall, a high engagement of students in creating their concept maps. This can be seen by the relatively high mean numbers of findings and differential diagnoses added to the maps in all pilots and the reasonable time students spent on the VPs. However, also here we see differences between pilots, which are especially visible in the mean number of connections created, which are highest in the pilots with students from Paris and Krakow, but quite low in the pilots implemented by Munich and Augsburg. Differences might be explained by the competency level of students, differences in instructions, experience with creating concept maps, or usability issues. Diagnostic accuracy depends on the difficulty and complexity of a VP. Some VPs have been piloted by different partners and also here we can see some differences. For example, the VP "Jelena Jakovic" was piloted in Krakow with 85.5% diagnostic accuracy vs.

25.0% in Latin America or the VP “Caroline Bach” with 83.5% diagnostic accuracy in Krakow vs. 69.8% in Paris. How difficult a VP is for a group of students depends on many factors, such as familiarity with the disease, level of competency, or prior knowledge. In a few VPs in different languages diagnostic accuracy was surprisingly low. For example, the Polish version of "Leslie Smith" was not diagnosed correctly by any student or the Spanish version of "Ulrike Birnbaum" was only solved correctly on the first attempt by 13.3% of participants. Such extreme data are valuable feedback for quality management. Based on the data we checked these VPs and made changes accordingly.

To further investigate such differences between VPs and countries, we see a need for future controlled studies, which also should include a thorough qualitative analysis of created maps. Interesting factors to control will be for example, gender, prior knowledge, and experience with concept maps of students, the influence of contextual factors in the VP, which we carefully designed and described in our blueprint, and the influence of the setting (asynchronous, synchronous, or blended) and the role of the teacher.

5. References

1. Huwendiek S, De Leng BA, Kononowicz AA, Kunzmann R, Muijtjens AMM, Van Der Vleuten CPM, Hoffmann GF, Tönshoff B, Dolmans D. Exploring the validity and reliability of a questionnaire for evaluating virtual patient design with a special emphasis on fostering clinical reasoning. Medical Teacher. 2015;37(8):775–82.
2. Huwendiek S, Haider J, Tönshoff B, De Leng B. Evaluation of curricula integration of virtual patients: Development of a student questionnaire and a reviewer checklist within the Electronic Virtual Patient (eViP) project. Bio-Algorithms and Med-Systems;2009;5:35-44.