

**DEPARTMENT OF LEARNING, INFORMATICS, MANAGEMENT
AND ETHICS**

Karolinska Institutet, Stockholm, Sweden

**Virtual Patient Simulation:
Implementation and use in assessment**

Mihaela Botezatu



Stockholm 2010

Published by Karolinska Institutet. Printed by US-AB, Stockholm, 2010.

© Mihaela Botezatu, 2010.

ISBN 978-91-7457-092-2

"It's a dangerous business, going out your door. You step onto the road, and if you don't keep your feet, there's no telling where you might be swept off to."

Bilbo Baggins

Dedicated to those who travelled with me.

ABSTRACT

Virtual Patient Simulation systems (VPS) are educational tools now considered to have entered the mainstream of medical education. VPS support not only undergraduate learning - where they are used mostly for learning and training clinical reasoning -, but for continuing medical education and patient orientation as well. Regardless of educational setting, the broad use of virtual patients for learning has not been paralleled by matching research efforts regarding implementation issues or the educational results of VPS use. The scope of the present research was therefore *i)* to highlight the must-have features of a VPS leading, in the eyes of different stakeholders, to a successful implementation of similar applications and *ii)* to clarify the educational results of VPS implementation for learning *and* assessment.

The results of the present studies convey the importance of several VPS features and educational uses, such as: end-user customization; authenticity of the software design, clinical scenarios, media used to support the case and case feedback; use of VPS for clinical reasoning development, in a broad curricular context of clinical specialties, supporting learning of topics not seen during clinical rotations; and a needed relevance of the VPS assessment for the future clinical practice. Assessment with VPS, arguably one of several components of the continuum of implementation, yields better results than ordinary course evaluation when the VPS applications are used both for learning and for assessment. Interestingly, delayed (long term) retention in VPS students also exceeds that of their peers exposed only to traditional learning and evaluation methods.

The findings also indicate that if virtual patients are to stay in the mainstream of medical education, developers, educators and researchers may soon have to deal with issues such as the continuum of VPS implementation, the authenticity of virtual patient design and clinical scenarios, as well as end-user customization. Accountability and sustainable development profile themselves as imperatives for the virtual patient simulation field.

Keywords: medical education, virtual patient simulation, implementation, assessment.

LIST OF PUBLICATIONS

- I. **Botezatu M**, Hult H, Tessma M, Fors U: "As time goes by: stakeholder opinions on the implementation and use of a virtual patient system". *Medical Teacher* 2010;32(11):509-516.
- II. **Botezatu M**, Hult H, Fors U: "Virtual Patient Simulation: what do students make of it? A focus group study". *BMC Medical Education*. Submitted.
- III. **Botezatu M**, Hult H, Tessma M, Fors U: "Virtual Patient Simulation for learning and assessment: superior results in comparison with regular course exams". *Medical Teacher* 2010;32(10):845-850.
- IV. **Botezatu M**, Hult H, Tessma M, Fors U: "Virtual Patient Simulation: knowledge gain or knowledge loss?" *Medical Teacher* 2010;32(7):562-8.

The publications listed above will be referred to as studies I - IV in the text.
Articles reprinted with permission from Informa Healthcare.

CONTENTS

ABSTRACT	i
LIST OF PUBLICATIONS	ii
INTRODUCTION	1
AIMS OF THE RESEARCH	8
METHODS	9
RESULTS	21
DISCUSSION	35
CONCLUSIONS	51
FINAL WORDS	53
ACKNOWLEDGEMENTS	54
REFERENCES	56
STUDIES I-IV	61

INTRODUCTION

I. Definition

Virtual Patient Simulation (VPS) systems can be defined as “interactive computer simulations of real-life clinical scenarios for the purpose of healthcare and medical training, education or assessment” [23] or as “computer programs that simulate real-life clinical scenarios in which the learner acts as healthcare professional obtaining a history and a physical exam and making diagnostic and therapeutic decisions” [33].

II. A bird’s - eye view of the field in 2010

Unlike their close relatives - the standardized patients and the high-fidelity simulators¹ - VPS systems are fairly recent additions to the spectrum of simulation. In 2009, Virtual Patients were considered to have finally entered the mainstream of medical education [25] – that is, 10-15 years later than standardized patients and about 5-10 years later than high-fidelity simulators. Early design flaws, use of improper educational models, high up-front costs and lack of proof of educational efficacy, together with “many other practical and cultural issues”, were among the reasons cited in the literature for such a late VPS acceptance [25].

Errors in implementation, absence of localization² and of post-implementation refinement, hesitations in the educational use and in curricular integration, combined with the lack of constructive alignment between learning and assessment, may have also taken a heavy toll on the perceived performance and utility of the VPS applications.

Despite a sense of ubiquity in healthcare sciences, it is difficult to ascertain how many VPS are used worldwide and for which educational purposes. From 142 US and Canadian medical schools requested in 2005 by Huang and colleagues to report on VPS activities, 26 of the 108 respondents were involved in VPS production, without mentioning the actual use of the application [33].

¹ Standardized patient: human actor trained to act as a real patient in order to simulate a medical disorder. High-fidelity simulator: computerized, interactive, life-sized manikin programmed to provide realistic physiological responses.

² Localization: VPS adaptation to the socio-cultural and medical context of the host institution / country.

A search of major databases using the criterion *virtual patient*³ was commissioned at the Library of Karolinska Institutet in May 2010. For the period 2000 - 2010, the search retrieved 228 articles indexed in Medline and 289 in the Web of Science; 153 (67%) were published between 2005 and 2010; 46 (20%) were published in 2009 alone. The abstracts of the 228 papers were reviewed in the light of the above-mentioned definitions and only 63 articles were found to be truly about VPS; the rest dealt with virtual worlds, virtual reality, computational modelling for drug delivery or radiotherapy administration, virtual simulations of surgical skills and so on. Of the 63 VPS articles, a mere 7 reported educational results of VPS applications, while 6 had connections with various implementation issues. Even if these absolute numbers may seem surprisingly low, they do make sense, considering that a 2006 review of “simulation in medical education” found just 13 papers on assessment, out of a pool of 232 articles published in 2005 [13].

III. Design and educational uses of VPS

In recent years, VPS have expanded their use for learning – which is mainly to foster the development of clinical reasoning skills, according to Cook’s continuum of competency [19] (Fig. 1).

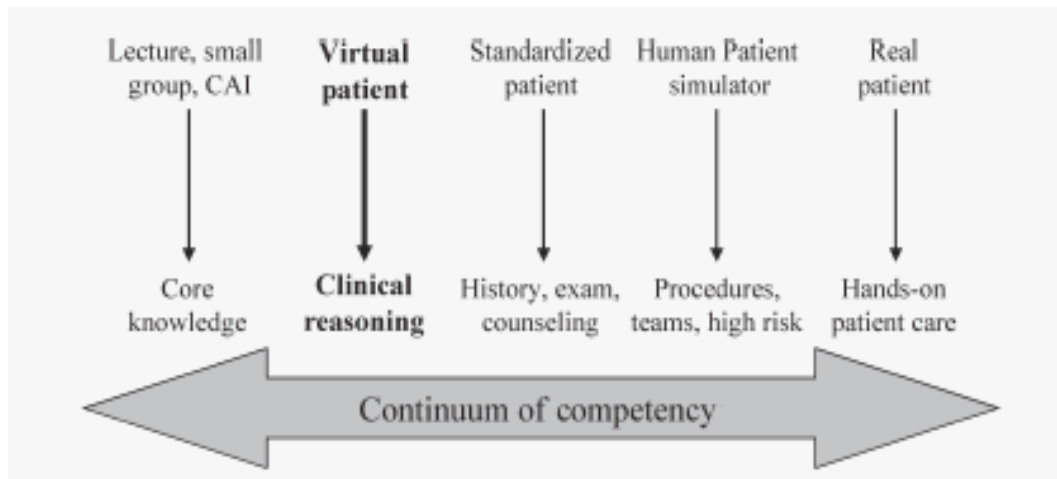


Figure 1. The aligning of instructional modalities with desired outcomes in a continuum of competency. CAI, computer-assisted instruction [19].

According to literature, VPS support not only undergraduate learning, but continuing medical education and patient instruction as well. Regardless of educational setting, the broad(er) VPS

³ Search queries different from “virtual patient” - such as “simulated patients” or “computer-assisted instruction” – found more articles, however not related to VPS. They were therefore discarded as search terms.

use for learning has not been paralleled by a purposeful participation in assessment. However, achieving constructive alignment (a concept that became one of the cornerstones of higher education soon after John Biggs coined the term in 1999) is extremely difficult. Rather than a rational top-down course design, a reflective practice is needed, constantly adapting the design of the course and the methods for its delivery (e.g. VPS) to the intended learning outcomes. When successful, such an approach will ensure that the learning and assessment activities are aligned with the proposed learning outcomes (e.g. in the case of VPS, the fostering of clinical reasoning).

The educational applications of VPS are intimately connected to their design [24]. To reach the proposed learning goals most systems take advantage of the linear-interactive, branching or knowledge-based contextualisation layout. For example, Web-SP (from Karolinska Institutet, Sweden) is a linear-interactive application [62], while Labyrinth (from the University of Edinburgh, UK) has a branching layout (<http://labyrinth.mvm.ed.ac.uk/>).

The users of VPS systems are typically required to perform a variety of tasks, depending on the educational objective (find, solve, prevent, explore, critique, experience, collaborate, diagnose, treat); medical students, however, suggest that the tasks be authentic, allowing them to make all the “decisions a real doctor would make” [34]. A description of typical user tasks, accompanied by screenshots from the Web-SP system, can be found in the Method section.

IV. The continuum of implementation and educational use of VPS systems

VPS, the beginning

The way VPS applications are “born” has profound consequences on their future implementation, since VPS development requires the careful examination of the educational goals of a given institution, combined with broad consultation among clinical experts, education specialists, IT developers and faculty administration. In this concerted-creation scenario, the initial development is followed by a preliminary instructional implementation in a course, normally restricted to a few course topics. This is because, no matter how brilliant, any VPS software still needs a handful of good quality clinical cases to begin with, and the creation of cases is a complicated matter (e.g. the literature has not yet demonstrated how many cases are enough for reaching a particular educational objective, how these cases should be collected / created or what makes a virtual case a “good” teaching case).

So developers at the original (“mother”) institution usually begin the creation of an application; however, the developing team may in some cases be restricted to IT staff, which in turn may lead to delayed acceptance of the product by clinicians and faculty, as well as to limited implementation – as a pilot or as an add-on to the already overcrowded medical curriculum –, a situation that will inevitably self-limit the use of the system. Regardless of its original composition, if the developing team cannot offer customization services according to the end-users’ needs (i.e. those of the clinical educators), the situation soon becomes an “arrest in maturation”: the application is restricted, for the entirety of its useful life, to the possibilities embedded in it during the creation phase.

The short, happy life of a VPS application

If everything goes well, the application is successfully implemented and begins its useful life in the given course. But what is the lifespan of a VPS? Difficult to say, as the literature is reticent in depicting this phase. What is known for sure is, for example, that the CLIPP project (reporting the use of the CASUS template for teaching the US core paediatrics curriculum), was ongoing between 2003 and 2005 [26]); Huwendiek and colleagues report the use of the CAMPUS system in Heidelberg from 2000 (in undergraduate paediatrics), but fail to reveal the date of their focus group interviews, in an article on VPS design published in 2009 [34].

Regardless of lifespan, which could be approximated at a minimum of two years, several interventions are required during the post-implementation setting: a continuous creation and implementation of peer-reviewed clinical cases (hopefully for constructively aligned learning *and* assessment), offering authentic learning scenarios for clinical reasoning development; a continuous update of the already implemented cases, according to the current standard of medical practice; an evaluation of the achievement of the educational goals with VPS, whenever deemed necessary; a permanent striving for curricular integration, staff development, research and publication, exchange of cases with other institutions, localization and customization. In other words, VPS demand high maintenance and a controlled implementation – post-implementation environment. Unfortunately, the obvious need for VPS sustainable development has not been substantiated by literature so far.

The mother - daughter relationship

Developing VPS is a notoriously resource-consuming endeavour [26, 33, 19, 52]. For such obvious reasons, many institutions choose not develop their own VPS application. Instead of novel VPS creation, the use of existing systems might prove advantageous [30]. Let's just consider again the example of the CLIPP project, where the CASUS proprietary software package was chosen for case authoring of 31 cases, in order to avoid the significant time and expense of software development.

A few pre-adoption steps applicable to VPS are described in the literature [16, 49], but bypassing preliminary stages is not uncommon practice during the adoption procedures. As a consequence, the teachers and students of the daughter institution may be minimally involved in the decision to adopt a specific application and as a result only reluctantly agree to test the latest educational "gadget". The lack of further development by the mother institution - mainly as localization to the host's socio-cultural and medical context and as customization according the specific needs of the end-users [28] - hinders the use of a successfully implemented application. Again, the sustainable development (i.e. localization and customization) of the adopted application, either by the mother or the daughter institution, is a necessary step towards fulfilling the educational scope of VPS implementation.

The stakeholders and the Issenberg product

In 2006, Barry Issenberg described three components essential for the effective use of any simulation: training resources, trained educators and curricular institutionalization. He postulated that "if any of these components are missing or deficient, the product will become zero and effective training will not occur" [36]. Conceived mainly with procedural simulation⁴ in mind, the Issenberg product can be applied to VPS as well, provided a few minor corrections are made. They should account for the stakeholders' opinions, which are in fact modelling the VPS patterns of implementation and use, but also for the needs, characteristics and educational context of the end-users. In spite of promising research results in recent years and of some attempts to integrate VPS in different curricula, the mentioned aspects have not yet been elucidated and a few years later it is still felt that "effective use requires evidence to guide design and integration" [19].

⁴ Procedural simulation: simulation modality aiming to improve actual operational performance in clinical settings.

The silence of the VPS

Virtual Patients probably fade away and die of neglect. Even if there is anecdotal evidence that a VPS is actually dead and buried in an unmarked grave, no research article has yet registered the death of a VPS. A pseudo-obituary is needed and should clearly mark the accountability of the developing team, of the origin / host institution and of the researchers themselves, towards the VPS end-users, in particular, and the medical community, in general. Accountability is quintessential to sustainable development. The academic community should be informed about less glamorous topics, such as arrested development, actual number of users, useful lifetime of the applications, direct and indirect costs of development and maintenance, localization efforts, staff development and implementation mistakes, or else the medical community will continue to be torn between the “band-aid / asset in the learning portfolio” disjunctive [59].

The implementation of a VPS application is a more encompassing concept than the name would suggest. Deeply, and sometimes irreversibly, influenced by the software development team, VPS implementation comprehends both initial implementation steps and post-implementation interventions - the latter, essential for the survival and the proper educational use of the application. A *continuum of implementation* therefore characterizes the entire lifespan of a virtual patient simulation system.

The research

The scope of the present research project was to clarify aspects crucial to VPS implementation and use. The research questions formulated here derived from my everyday practice with a variant of Web-SP (Web-based Simulation of Patients) at my home university in Bogota, Colombia.

In 2005, Web-SP - an application developed at the Department of Learning, Informatics, Management and Ethics, Karolinska Institutet - was implemented at Universidad el Bosque, Bogota. The original application was translated into Spanish and adapted to the local medical practice and socio-cultural perspective of Colombian doctors and patients. The resulting Spanish variant of Web-SP was integrated in the Internal Medicine curriculum of the Faculty of Medicine, by decision of the University Board. The cases serving for virtual patient creation were collected locally, from four university hospitals, by internal medicine students; a pool of more than 200 real-life clinical records with accompanying media were available by

the end of the project. For the scope of this research project, virtual cases were developed for two topics in Internal Medicine (namely haematology and cardiology), as the respective clinical teachers expressed their wish to create cases and use them before their colleagues from other sub-specialties.

It is only fair to say that the implementation process and the use of the application in the classroom for teaching and assessment raised many questions among students, clinical teachers and board members. Answers to some of them (i.e. regarding VPS implementation and their use for classroom assessment) can hopefully be found in the following pages. Two of the studies are directly dedicated to general implementation issues, while the other two explore more closely a component of the continuum of implementation, the use of VPS for assessment purposes.

AIMS OF THE RESEARCH

The overall aim of the research project was to increase our understanding of the implementation and the educational use of Virtual Patient Simulation systems.

The specific research questions addressed were the following:

- According to different stakeholders, which VPS features contribute most to a successful implementation and use of such applications?
- What is the opinion of medical students on the educational use of VPS?
- Is there a difference in evaluation results between the virtual patient application and the regular course examination?
- Can a VPS system, such as Web-SP, contribute to superior delayed (long-term) retention of knowledge in a clinical course?

METHODS

The study population

Between 2006 and 2008, a cohort of undergraduate medical students (n= 216) from the Faculty of Medicine, Universidad el Bosque (Bogota, Colombia), participated in the four studies. All the students were enrolled in the Internal Medicine course, during four consecutive terms. According to the local curriculum, it was their first contact with clinical sciences. It was estimated that the participants possessed a similar body of knowledge and skills and that their IT knowledge was homogenous [5]. The mean age of the participants ranged from 21.7 to 22.7 years [5].

Besides medical students, faculty board members and university teachers from both Colombia and Sweden (Karolinska Institutet and Halmstad University) agreed to participate in study I.



Photo 1. A “sample” of the study population at Universidad el Bosque.

The curriculum

At Universidad el Bosque, Internal Medicine is a five months’ course placed at the beginning of the fourth year (out of six years in total), right after preclinical studies. An intensive initial module of lectures (on average, five topics per sub-speciality, haematology and cardiology among them), small-group assignments centred on solving clinical cases, combined with a variety of procedural simulations, is followed by clinical rotations in internal medicine wards

at university hospitals in Bogota. Course assessment is scheduled at different time-points during each term. In March 2005, the VPS system Web-SP became part of the Internal Medicine curriculum at El Bosque Medical School; as a consequence, all the students enrolled in the Internal Medicine course (45 - 65 per term) were recruited into the studies. Curricular integration began with, but was not limited to, the cases developed for the scope of the trials.

The application

Web-SP (Web-based Simulation of Patients) - an explorative, linear-interactive virtual patient simulation developed at LIME, Karolinska Institutet - was the virtual patient application used throughout the research [62]. A Spanish variant of the software was developed and used for teaching and performance-based assessment at Universidad el Bosque [28].



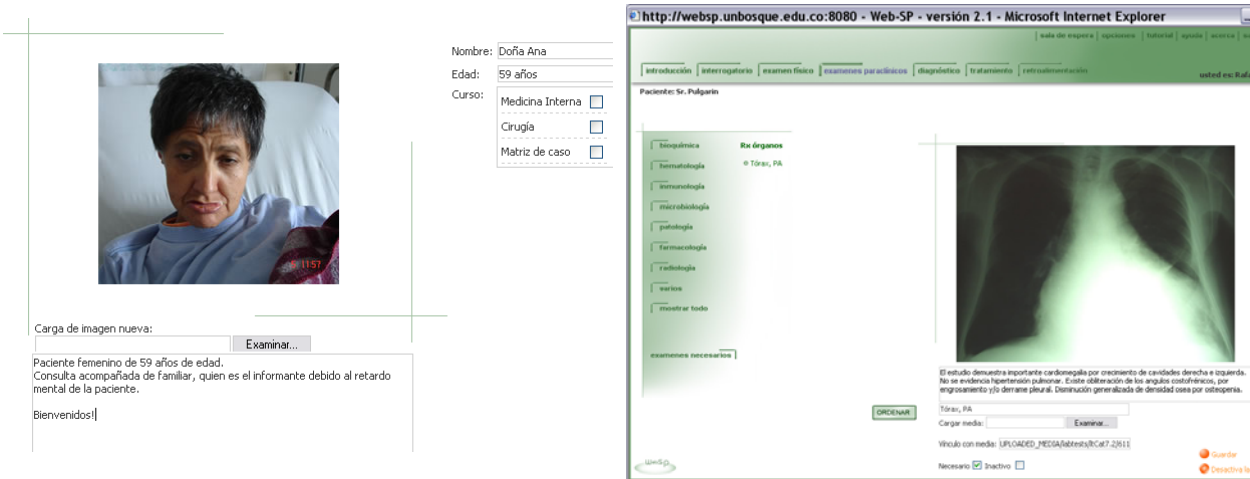
Screenshot 1. A view of the Internal Medicine module in Web-SP-Español.

At the beginning of each term the participants were introduced to Web-SP features by means of a live presentation and of a virtual patient case demonstration. A Web-SP tutorial was uploaded on the *moodle* e-learning platform used by teachers and students to support learning and assessment during the course.



Screenshot 2. Web-SP-Español support features on moodle.

In order to solve a virtual patient case in Web-SP, the students were required to obtain a medical history, obtain and interpret physical exams, order appropriate laboratory tests and finally formulate a diagnosis and treatment course, both entered as free-text responses.



Screenshot 3. Cases in Web-SP-Español.

At the end of a typical 45-60 minute session, the students could access the feedback module.



Diagnóstico sugerido
Leucemia-linfoma de células T-precursoras.

Complicaciones:
1. Derrame paraneoplásico.
2. Síndrome de lisis tumoral.
3. Síndrome de dificultad respiratoria aguda.

Explicación
Positividad para: CD3 de superficie, CD99, Tdt (en todos los casos); CD7; expresión variable de otros markers de células T cell markers: 50 % CD38, 33 % CD10; 20 % CD16, CD57 (markers de linfocitos NK); CD71 (receptor para transferrina).
Positividad según estado:
Estado I Pan T+, CD4+, CD8-, CD1-
Estado II CD1+; CD4 y CD8: o ambos positivos, o ambos negativos
Estado III CD1-; CD4+, CD8- o CD4-, CD8+



El tratamiento recomendado
1. CHOP (high-dose)
2. Irradiación local para masas tumorales grandes
3. Terapia del síndrome de lisis tumoral
4. Profilaxis del SNC
5. TMO

CHOP: ciclofosfamida (cytotoxic); adriamicina (doxorubicina); oncovin (vincristina), prednisolona. 4 ciclos cada 21 días; otros 4 ciclos si el paciente no responde/responde parcialmente.

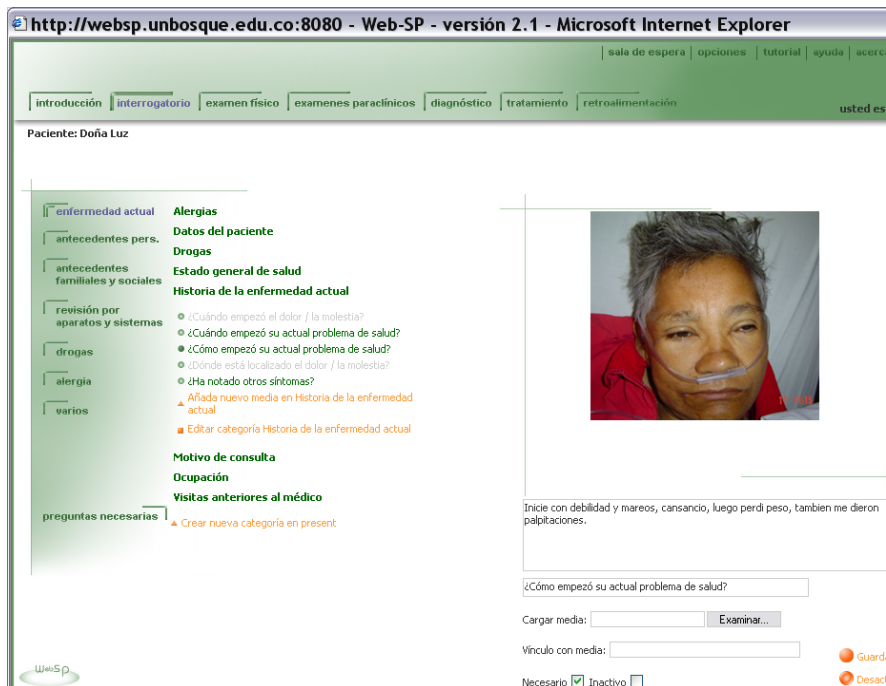
Ciclofosfamida, 750 mg/m², iv, en 15 min, Día 1
Doxorubicina, 50 mg/m², iv, bolus (1-2min), Día 1
Vincristina, 2 mg, iv (1-2 min), Día 1
Prednisolona, 100 mg, vo, Días 1 a 5

Efectos secundarios:
1. pancitopenia

Screenshot 4. The feedback module in Web-SP-Español. The actual feedback is longer than shown in the picture.

The cases

All the cases used for learning and for assessment in this research - both as virtual patient and as paper cases - were developed by senior clinicians from real clinical records collected locally. The virtual cases featured patient photos and diagnostic media pertaining to the actual clinical records.



Screenshot 5. A patient's face photo in Web-SP-Español.

All the patients whose clinical records served as a base for case development had signed an informed consent form, enabling the case creators to use data from the clinical history and media related to the case; data collection was approved in writing by the respective hospitals. The virtual patient cases were peer-reviewed before implementation. As for the studies on assessment, the contents of both Web-SP and paper cases were matched for topic and difficulty level. The feedback module in Web-SP provided an extensive clinical expert discussion of the case and also the actual clinical evolution of the real patient.

Synopsis of methods

Research topic	VPS implementation		Assessment with VPS	
Method	Qualitative		Quantitative	
Methodology	Semi - structured questionnaires	Focus-group interviews	Randomized controlled trial	
Study	Study I	Study II	Study III	Study IV
Population	n = 39 (MS, FB, UT)	n = 16 MS	n =216 MS	n = 49 MS

Table 1. Synopsis of methods. MS, medical students; FB, faculty board members; UT, university teachers.

Details for studies I - II

The students involved in studies I and II (n=16) belonged to term 3 (n=49) of the general population of undergraduate medical students. Term 3 was the one cohort geographically available at the time of both studies (convenience sampling).

Besides medical students, faculty board members and university teachers from Colombia and Sweden agreed to participate in study I. In Colombia, all five faculty board members involved in decision-making around the application and the 7 university teachers who used Web-SP in teaching and assessment at El Bosque Medical School were invited to answer the questionnaire. In Sweden, 3 faculty board members and 8 university teachers from Karolinska Institutet and from Halmstad University answered the survey (out of a total of 30 Web-SP users); no students were available in Sweden for the purposes of study I. Web-SP use in the two Swedish universities was optional at the date of the survey and encompassed medicine, nursing and dentistry programs.

In study II, the 16 medical students were allocated by simple randomization to one of the two focus group interview sessions. Randomization was performed to ensure representativeness, so that students with "good experiences" with Web-SP in terms of assessment results (who could have been more positive about the educational use of the application) were not purposely selected. We deliberately recruited more students than needed, in order to allow students to decline participation or drop out later. The randomly chosen students were directly contacted by the vice dean's office and invited to participate in the study. None of the students declined participation, nor dropped out of the study. When all students confirmed their participation, we arranged two focus groups, as 16 participants would have been too many to accommodate in one session.

Both studies on implementation relied on a qualitative-methods approach.



Photo 2. A focus group interview.

Study I.

A semi-structured questionnaire was developed for the purposes of study I. Two experts in medical education, co-authors of study I, reviewed the questionnaire for content validity. The research questions, divided into four themes of seven items each, were explored both quantitatively and qualitatively, using a mixed-method approach. The participants were asked to rank several aspects related to the implementation and use of Web-SP according to their perceived order of importance (on a 1 to 7 scale, from least to most important), while a number of open-ended questions were meant to further clarify issues raised as closed questions in each theme. The questionnaire explored items related to *a)* the implementation process; *b)* post-implementation activities; *c)* the curricular use of the application in teaching

and assessment, including desired features in support of those functions; and *d*) the possibility to extend the use to educational settings beyond undergraduate courses. The survey was conducted in Spanish, at Universidad el Bosque, where the instrument was subject to prior on-site peer validation before application; in Sweden, the chosen language was English and no validation was deemed necessary. No further actions were taken by the research team in order to address possible variations between the two languages.

Statistical analysis. Descriptive analysis was used to determine the median score and the inter-quartile range. Friedman's statistic was performed to test the hypothesis that there was no systematic difference in ranking the order of importance among the respondents. Kendall's *W* test was also applied to measure the level of agreement in the ranking of ordering of the relative importance of the items. The Jonckheere –Terpstra test was employed to test the null hypothesis that the medians were the same and Friedman's posthoc test was used for multiple non-parametric pair-wise comparison. The significance level was specified at 0.05. The statistical analyses were performed using SPSS and Statistica 8.0 software; the graphical presentations were created by Statistica 8.0.

Analysis of the open-ended questions. Since the open-ended questions were naturally pertaining to a theme and a sub-theme, and the answers had been brief, no formal content analysis was undertaken. All the opinions expressed were registered.

Study II.

Setting. The interviews were conducted in the facilities of the Faculty of Medicine. The author led the focus group interviews, assisted by one of the local clinical teachers. The discussion began by recapitulating the participants' experience with the system and by stating the scope of the interviews, i.e., the exploration of their perception of the application. Then several topics were introduced, as the students had used Web-SP for learning and/or assessment and had also collected clinical cases in the hospitals. During the interviews the students raised issues the researcher did not intend to bring up, such as communication skills development or their motivation in using the system.

Data collection. The interviews were conducted in Spanish, by the main investigator. An interview guide had previously been developed and peer-validated on site. Both audio and video recordings were made during the interviews, which were 60 and 62 minutes long, respectively. No field notes were taken during or after the focus group interviews. Each topic was discussed till estimated saturation.

Data analysis. The audio tapes were transcribed by the main author. A non-verbatim approach was used, meaning that the transcriptions were edited for pauses, interjections and other audio utterances, whose omission would not alter in any way the message conveyed by the participants. The video recordings were used to identify the speakers on the audiotapes, which were assigned consecutive numbers. The transcripts were translated into English by an authorised translator not affiliated to Karolinska Institutet or to Universidad el Bosque. The theoretical framework used in this study was content analysis [42, 43]. MB and UF reviewed the transcripts independently, compared notes and reconciled the differences. Emerging coding was used to obtain categories and themes, whose final form was reached after another consensus discussion between the two coders. Given the manageable length of the two focus group discussions, no software was used to analyze the data.

Details for studies III - IV

The studies on assessment explored the educational results of Web-SP use by means of randomized controlled trials. The whole student population in the medical program at El Bosque (n=216) participated in study III, while only students belonging to the third term (n=49) were enrolled in study IV. After introduction to Web-SP, the participants were randomized 1:1 to the Web-SP group (study group) or to the non-Web-SP group (control group), by means of a computer-generated randomization list. The first three weeks of each term were dedicated to the learning component. Two sub-specialties, haematology and cardiology, were chosen from the main course plan for the creation of virtual patient cases. For the purposes of the studies, the learning component for haematology and cardiology was comprised uniquely of Web-SP cases for the study group, while the controls had access only to the more traditional teaching methods (lectures and small-group discussions), covering the same topics. The students worked through the cases individually and completed them in approximately 45-60 minutes.

In study III, three different interventions took place during the learning component (Fig. 2). In term 1 and 2, the study group gained access to Web-SP cases, while only more traditional teaching-learning methods (lectures, small-group discussions and mannequin simulations) were available for the control group. In term 3, the study group had the benefit of both virtual cases and traditional teaching-learning methods (including lectures), while controls again had access only to the latter. For term 4, a paired design was used, enabling the students to be their own controls.

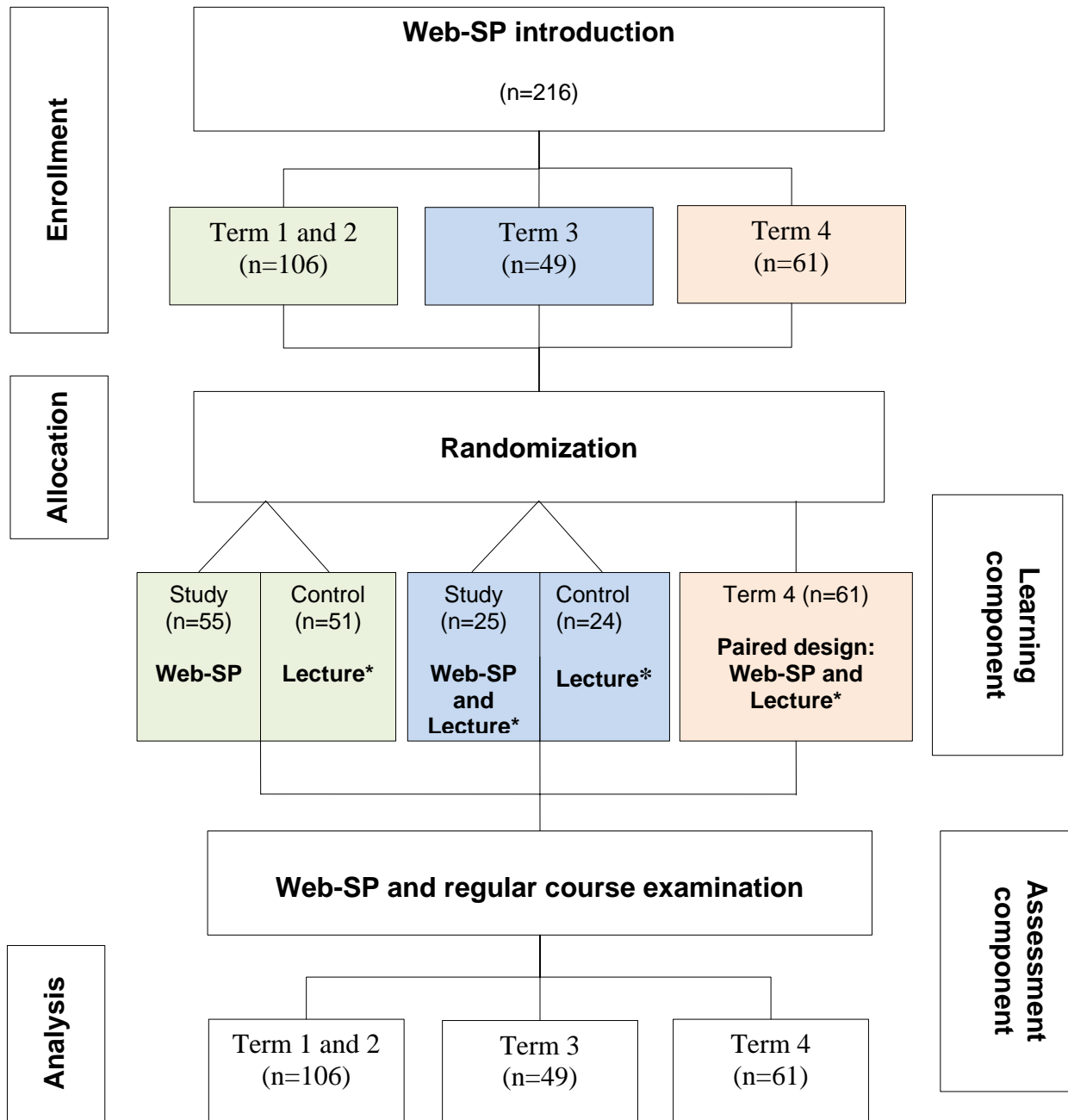


Figure 2. Flow of participants through each phase of the study III. * Lecture: includes teaching methods other than Web-SP.

In study IV, the initial learning component was followed by early assessment, as previously described; the participants were tested again at the end of the term, approximately 4 months after the first round of evaluation (Fig. 3).

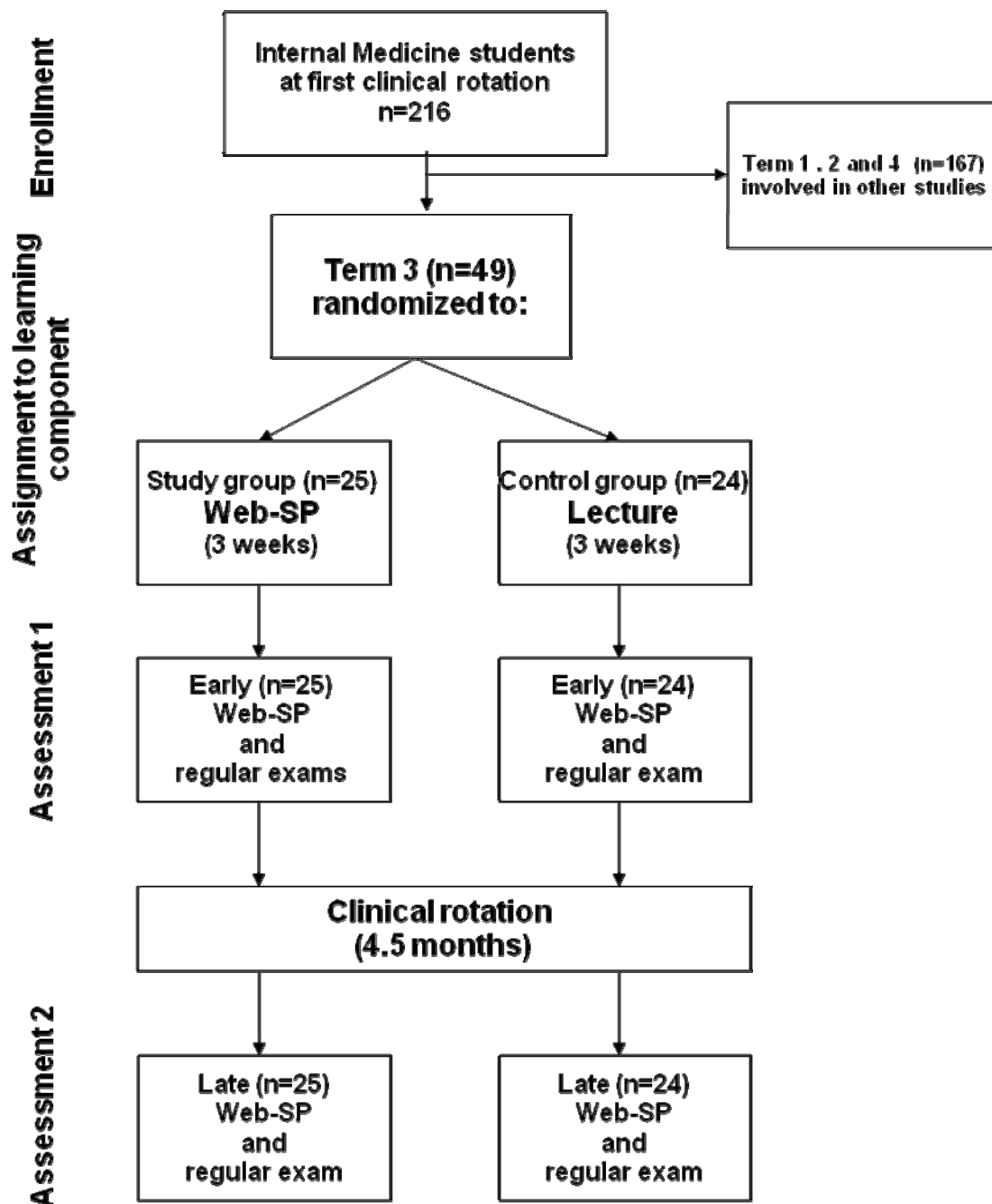


Figure 3. Flow of participants through the different phases of the study IV.

Regardless of randomization, the assessment component was unified throughout the studies and consisted of both virtual patient and paper cases (the latter, consistent with the evaluation methodology used in the Internal Medicine course before implementing Web-SP). Both forms of examinations were “production tests”, requiring the students to construct a response. In practical terms, the exam consisted of a VP haematology / cardiology case, followed by a paper haematology / cardiology case. The virtual patient and the regular exam cases were carefully matched for topic and difficulty level by a senior clinical expert.

A scoring rubric addressing the clinical reasoning process in items pertaining e.g. to diagnosis and treatment was developed for the purposes of the studies and applied to both assessment modalities. Positive points were awarded for “correct” decisions, while penalty points were subtracted in case of potentially harmful choices. To assess the reliability of the instrument, an inter-rater variability study was performed on similar cases at Universidad el Bosque in January 2008. In study III, the grading scale ranged from 0 to 6 points, distributed between diagnostic and therapeutic decisions; a more detailed scoring system, on a 10 point scale, was used in study IV (Table 2).

Item	Positive score	Negative score	Max. score
Patient interview			1
	only the options indicated as necessary by the case creator were assessed	+1	none
Physical examination			1
	only the options indicated as necessary by the case creator were assessed	+1	critical findings missing -0.5
Labs & imaging			2
	all options indicated by the case creator and no more than 10% unnecessary	+2	lack of 20% of indicated or more than 20% unnecessary -0.5
			serious consequences if missed -1
Diagnosis			3
	positive diagnosis	+1	potentially life-threatening decision -2
	co-morbid associations identified	+1	
	differential diagnosis with adequate discussion	+1	
Treatment			3
	adequate treatment for positive diagnosis	+1	potentially life-threatening decision -2
	adequate treatment for co-morbid associations	+1	
	lifestyle changes, diet, follow-up, referral to specialist, explaining surgery and prognostic, etc.	+1	
Total			10

Table 2. The scoring scale used in study IV.

The clinical rotations included formal cardiology rotations, while the instruction in haematology was opportunistic, as sporadic cases arrived at the internal medicine departments of each hospital.

Statistical analysis. In both studies, an independent sample t-test was employed to compare the means of the study and control groups, testing the null hypothesis (i.e., that the means were equal). The Levene's test of homogeneity of variance was computed to test the assumption of equal variance. The paired t-test was used to test the null hypothesis, that there was no difference in mean scores between the assessment with Web-SP and regular examinations / between the early and delayed assessments, respectively. The level of statistical significance was specified at 0.05 for all significant tests.

In study III, the internal consistency of the items was estimated by the Cronbach's-alpha coefficient. The inter-rater reliability was measured by the intra-class correlation. Tukey's test of non-additivity was performed to examine multiplicative interaction.

In study IV, the Mann-Whitney test was applied when distribution deviated from normality and also to compare the mean ranks of the two groups. The mean rank was used to ascertain which group had the highest score in the four categories (patient interview, laboratory, diagnosis and treatment). The effect size was calculated based on the means and the standard deviations.

RESULTS

Study I

The respondents did not show equal preference in rating the ranking of the order of importance of the variables (Friedman's Chi square: 26.5 to 115.1, $df = 6$, $P < 0.001$). Kendall's coefficient of concordance ranged from .11 to .50, reflecting that the ranking of ordering and level of agreement on the relative importance of the items is different among the subjects (Table 3). The range indicates a lesser degree of unanimity among the various responses. Kendall's W can be interpreted as a coefficient of agreement or disagreement among raters. The coefficient W ranges from 0 to 1, with 1 indicating complete inter-rater agreement, and 0 indicating complete disagreement among respondents.

Variable	Friedman's Chi-square	df	P-value	Kendall's W *
Implementation process	78.3	6	< 0.001	.34
Post-implementation activities	26.5	6	< 0.001	.11
Intended curricular use	107.1	6	< 0.001	.46
Use in different educational settings	115.1	6	< 0.001	.50

Table 3. Friedman's test and Kendall's W test by variable group (n=39).

* Kendall's coefficient of concordance

The Jonckheere –Terpstra test identified a significant between-group difference in some of the items for three major variables, namely implementation process, faculty post-implementation activities and curricular use, but failed to show statistically significant difference for “use in different educational settings”. However, significant within-group differences were observed in all variables. The median score of the ratings with the inter-quartile range is presented in the graphs (Figures 4-7). Significant between-country differences in the rating of the four major variables were not observed ($P > 0.05$).

The implementation process. A statistically significant between-group difference was not observed for the variable “implementation process” in all the items, except for “cost of the initial implementation” ($P < 0.001$). The within-group difference showed a statistically significant difference ($P < 0.001$), indicating that respondents have significantly different ratings for different sub-categories of the implementation process (Fig.4). The different participant categories agreed most on the importance of easy VPS customization by the end-user; a good design of the software, together with curricular integration, also ranked high in their preferences, regardless of stakeholder group.

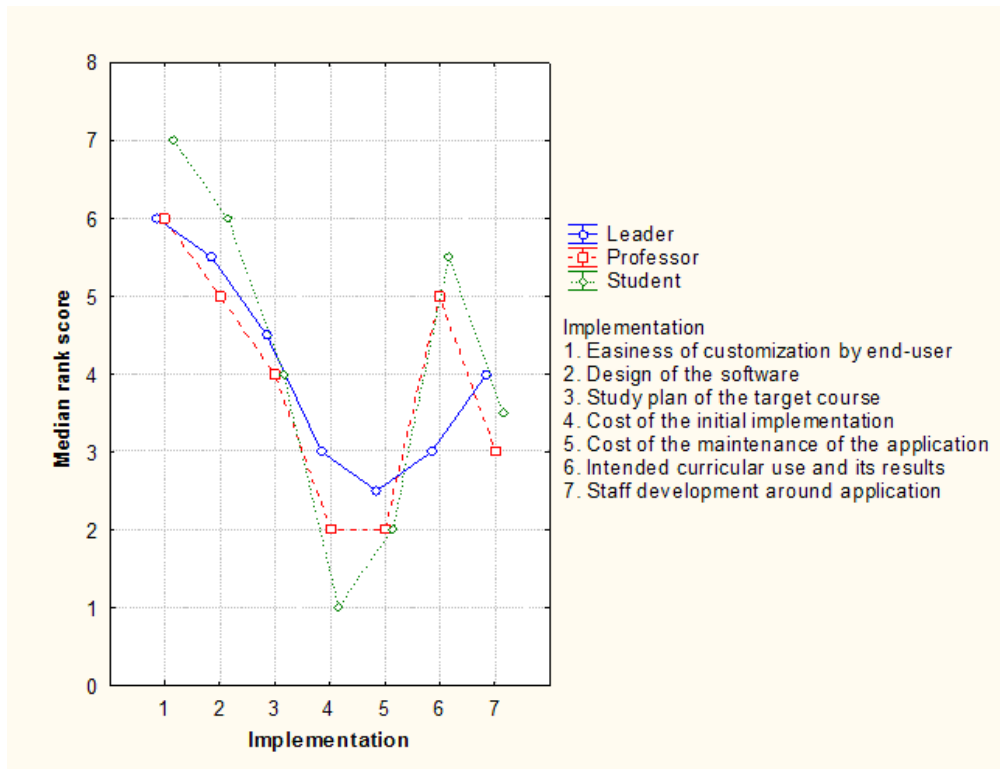


Figure 4. Median rank score of responses for the variable “implementation process”.

In the open-ended questions, the participants were asked about Web-SP features that needed improvement as soon as possible. The staff directly involved in case creation commented on the need for “deep” customization (beyond the actual capabilities of Web-SP), and some pointed out the importance of further development by the institution of origin. A number of students showed preference for branched cases, instead of the linear design of Web-SP, and indicated that such a flow would better support the clinical reasoning development. Teachers and students objected to the unrealistic design of the physical examination section, which detracts from the feeling of authenticity and makes case creation cumbersome and time-consuming. Both professors and students expressed interest in grouping the lab section in test “batteries”, instead of alphabetical listings, and discussed the real life improbability of ordering an unlimited number of tests on a single patient. The use of VPS as an add-on to an existing curriculum was strongly discouraged by participants from both countries.

Post-implementation activities to reach the proposed educational goals. No statistically significant between-group differences were observed in five of the seven sub-categories. However, significant between-group difference was observed in the “exchange of items with other HEI” ($P = 0.006$) and “staff development” ($P=0.02$), items which registered most disagreement among different stakeholder groups. The within-group difference was

significant. The highest level of agreement was observed around the necessity of the continuous creation of new cases. Figure 5 displays the between- and within-group differences.

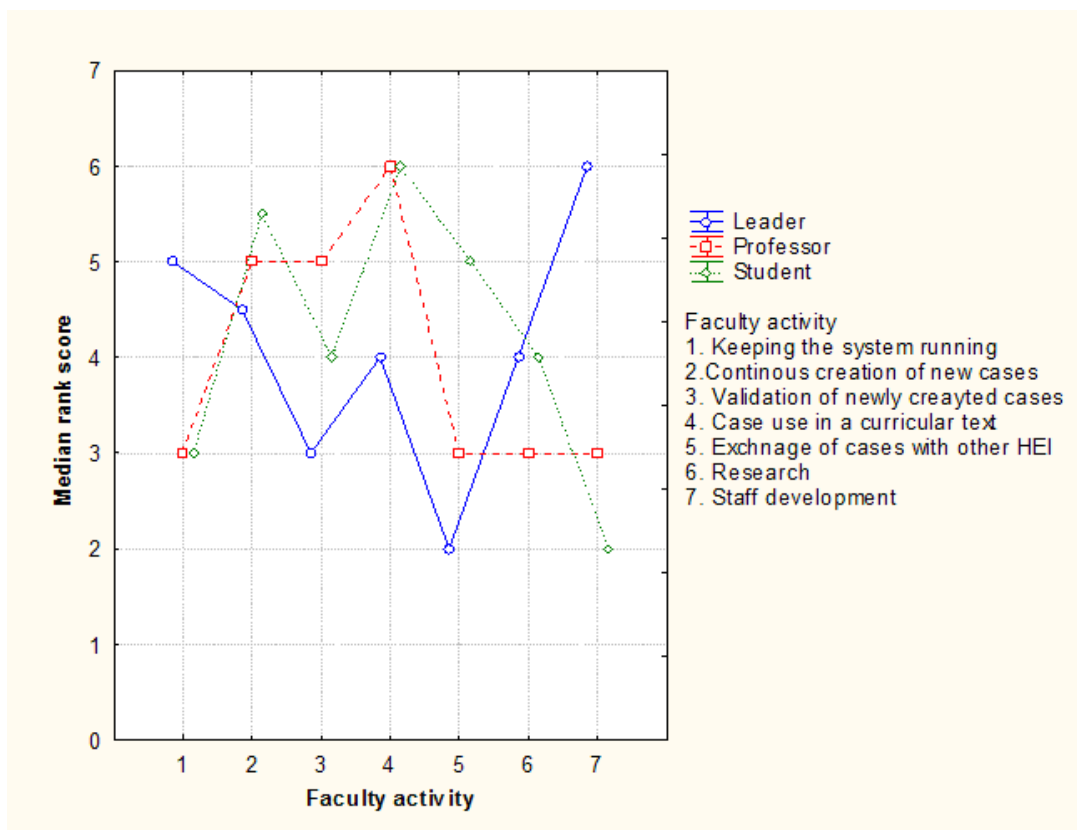


Figure 5. Median rank score of responses for the variable “post-implementation activities”.

The participants discussed items they considered essential to post-implementation. One of the most debated was the creation of new cases. The students called for numerous cases within a sub-speciality, with a minimum of 2-3 cases per disease. The professors also highlighted the importance of a “critical mass” of cases, but underlined that creating numerous cases is time-consuming. Most teachers and students believed that real-life cases are intrinsically better than fictitious cases.

Intended curricular use. The between-group difference for the teaching-learning component was not statistically significant, except for the “factual knowledge” ($P = 0.006$) and “topics not covered by the study plan” ($P = 0.04$) items. The development of clinical reasoning constitutes the main curricular use of VPS for all respondent categories, followed closely by VPS depicting unusual diseases and by common diseases in usual presentations, respectively. The within-group difference in rating the different items (Fig. 6) was statistically significant ($P < 0.001$).

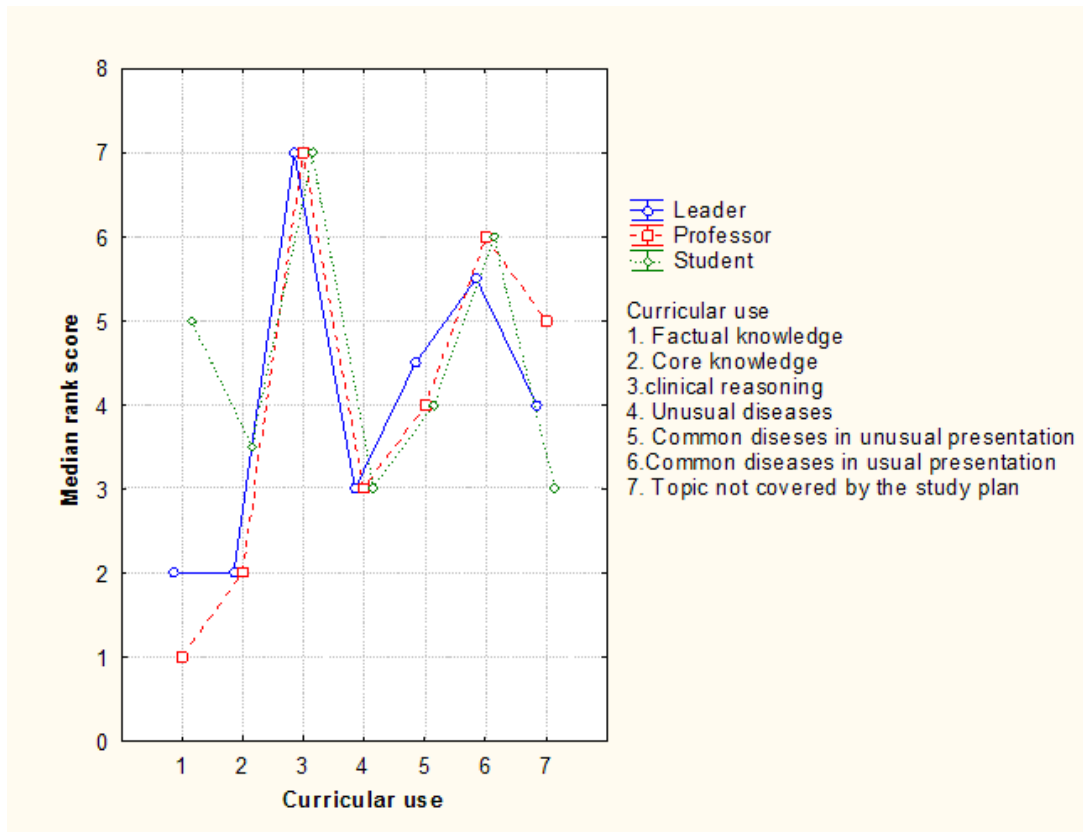


Figure 6. Median rank score of responses for the variable “intended curricular use”.

The stakeholders indicated the development of clinical reasoning as the most important use of VPS for teaching purposes, while core knowledge, and especially factual knowledge, ranked low in the professors’ and leaders’ perspective. Common diseases in usual presentations, followed by common diseases already complicated at diagnosis or in unusual presentations were considered to make a good curricular use for VPS.

The assessment component was addressed as open-ended questions only, referring generally to VPS. The participants strongly believed that VPS systems should not be the only assessment tool used in a course. They thought that VPS should not be used for high-stakes assessment either. All stakeholder groups encouraged VPS use in ordinary course assessment, in combination with other evaluation methods.

Use in different educational settings. The between-group difference was not statistically significant ($P > 0.05$) for all items. The within-group difference in rating the different items was statistically significant ($P < 0.001$). Significant levels of agreement were registered among the responder groups regarding the need for adapting the difficulty level to the educational setting where the VPS are used (undergraduate courses, postgraduate studies or continuing

medical education), followed by changes in design (i.e. a different flow through the application). See Fig. 7.

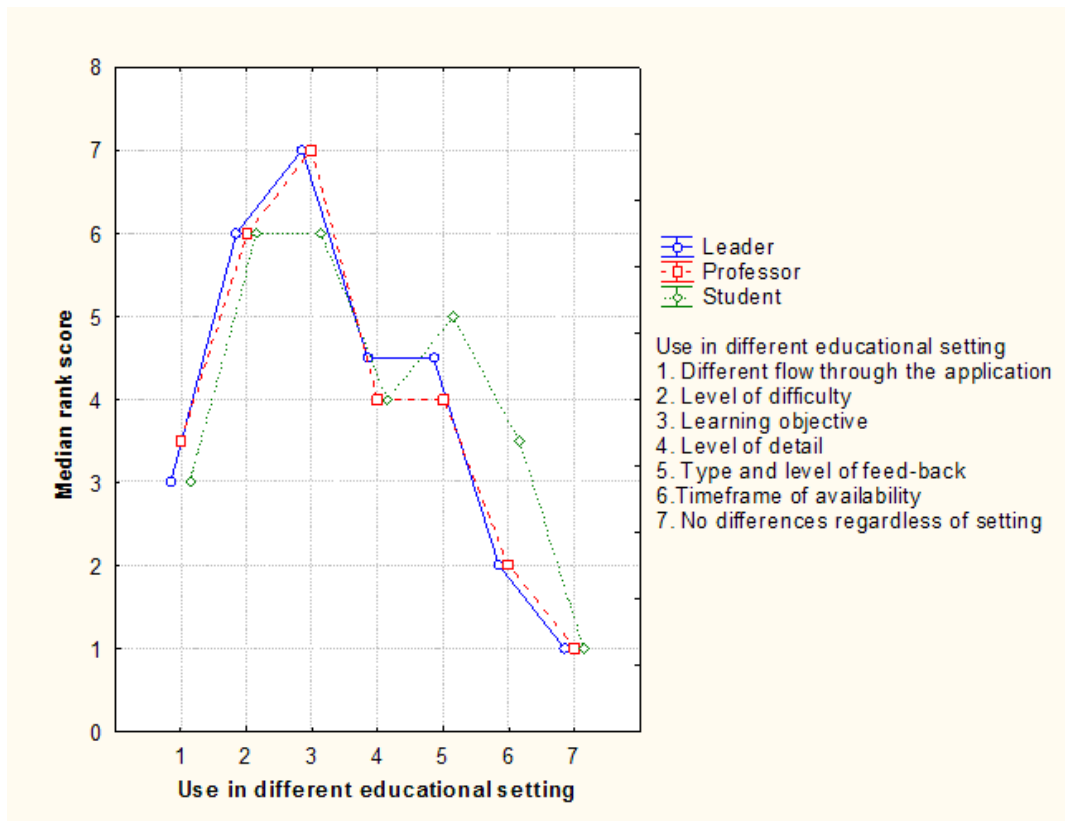


Figure 7. Median rank score of responses for the variable “use in a different educational setting”.

The various stakeholders were unanimous in indicating that learning objectives must be specific to the educational setting VPS are used for, closely followed by a different difficulty level and a different type of feedback. Most free comments pointed out that a case created for medical students is unlikely to be suitable for medical residents or for continuing medical education in nursing, for example.

Study II

Emerging coding of the transcripts identified eighteen categories, which were further clustered into five themes⁵ (*learning, teaching, assessment, authenticity, implementation*). The presentation of the results includes quotations, followed by the student identification number. The most articulate quotation was chosen in each case, in order to avoid redundancy. When several participants concurred, the identification numbers were omitted.

⁵ For student quotations, please refer to the original article.

Learning

- *Clinical reasoning.* The main educational scope of virtual patients found in the interviews is the use for training clinical reasoning, for which the constant use of VPS is deemed essential. Training with VPS is also thought to enhance analytic and synthetic reasoning. Clinical reasoning development is linked to a stepwise approach to case-solving, fostered by VPS. Even though clinical reasoning development is seen as the primary scope of learning with VPS, input on factual and core knowledge is welcome as part of the VP "package", especially as feedback; the VP scope "*would be to reinforce clinical reasoning abilities, but also to give feedback with regard to the theoretical aspects of the disease in question*". The participants greatly appreciate that VPS enable them to follow up a patient from the beginning to the end; VP cases give students a sense of closure, which enhances motivation. The students understand VPS as a "*preparation for real life as doctors*".

- *Transferable skills.* The real cases, seen later on as a young doctor, could be solved by association and comparison with virtual patient cases with similar characteristics. The knowledge acquired with VPS is considered transferable to other types of exams. Moreover, most students consider that knowledge transfers directly to the real patients, especially when the cases used in teaching were created from real life clinical records.

- *Retention enhancement.* When compared to other learning tools, VPS help best to retain information.

- *Making mistakes and learning from them.* Recognizing and correcting mistakes, either in the clinical reasoning path or in previously acquired knowledge, is considered crucial for successful learning. Moreover, students even think that they are less prone to repeat the mistakes made with VP in their clinical practice.

Teaching

- *Clinical specialties.* The participants believe that VPS should enjoy a broad use across clinical specialties. The majority of students favour the curricular use of VPS across the main clinical specialties, such as internal medicine, surgery, obstetrics-gynaecology and paediatrics. Within a given clinical specialty, the virtual patient cases should present frequent diseases and their complications; topics not included in the study plan and in the clinical rotations should also benefit from VPS, which could close the gap. Diseases that might easily be missed during a short clinical rotation (due to seasonality or to being endemic in a different geographic area), but that are relevant for future clinical practice, should also be presented as VPS.

- *Regulatory effect of VPS.* A possible regulatory effect of VPS is envisaged by the participants. Different rotation locations in Colombia offer a varied infectious disease spectrum, as well as patients with different socio-economic status and VPS are seen by most students as a chance to make instruction uniform among rotation sites.

The majority of students in our study seem to require external regulation, by means of limiting the availability of the system. Time restrictions were thought necessary, in order to help students to plan and organize their workload. The use of a system permanently left open is estimated by the participants at a surprisingly low 3%.

Assessment

- *Qualitatively different assessment with VPS than with regular exams.* According to the interviewees, VPS evaluate knowledge in a different manner, emphasizing the clinical reasoning process. The students consider VPS to be a more didactic form of evaluation and an intrinsically better evaluation tool than traditional exams. However, students agree that VPS should not be the only assessment form used in a course. As pointed out earlier, VPS also allow an increased retrieval of information in comparison with regular examinations and therefore grades tend to be higher with virtual patient evaluation.

- *Increased motivation.* VPS evaluation "*does not feel like an exam in the long run*". For example, Web-SP features open questions, which are "harder" than closed / multiple-choice questions; despite that, they "*feel more natural*" and the evaluation is deemed by most students as less stressful than with other assessment methods.

- *Professional focus in assessment.* Most students want to learn from assessment and from the feedback on assessment. Participants say they "*get more*" from VPS assessment. However, assessment should always be relevant for future clinical practice as a young doctor. VPS assessment, as well as evaluation in general terms, should be directed to "*what a general practitioner should know on a subject*".

- *Production assessment.* The students are positive about open questions, which "*make you think and analyze the patient*". They seem to favour open questions, even if the grades might be lower than with other types of evaluation.

Authenticity

- *Design and content.* In the opinion of participants, VP design should reflect real clinical practice. The option menus should offer realistic and localized choices in terms of history

taking, physical exam and diagnostic tools, as well as feedback on treatment alternatives. Realism could encompass even the actual costs for the healthcare system. Too many unstructured options in a VPS may be misleading; abundance is not necessary to meet the learning objective.

-Localization of the socio-cultural context. In the case of applications "born" in one country and "adopted" in another, it is crucial to adapt the patient interview section to the socio-cultural context of the patient.

-Realism and virtuality. The participants think that real life records make better cases than fictitious scenarios; moreover, such cases are thought to provide transferable skills. Media presence is essential to authenticity, and realism starts with the patient photo. A video recording is a must in certain circumstances. The participants no longer perceive the cases created from real-life patient records as "virtual".

-Feedback. Feedback on the actual evolution of the patient adds to the realism of the application. The feedback section should show the actual clinical evolution and the effect of treatment.

Implementation

A number of important issues were brought up by the students, including:

-Number of cases. More cases were linked to more knowledge. More than one case per topic can be necessary for common diseases which are often complicated / have co-morbidities at presentation.

-Access / availability. Broad access was preferred in relation to site - campus, hospital, home - , but not in relation to timeframe, which should be restricted, as previously shown.

-Virtual Patient exchange. The tropical diseases cases developed at Universidad el Bosque should be subject to exchange with other Higher Education Institutions.

For a clearer understanding of the results, a synopsis is provided below (Table 4).

Theme	Category	Main results
Learning	Clinical reasoning	-Clinical reasoning development is linked to a stepwise approach to case solving. -Input on factual and core knowledge is welcome as part of the VP "package", especially as feedback. -Holistic view of the patient and closure sense.
	Transferable skills	-Directly to the real patients, especially when the cases used in teaching were created from real life clinical records. -To other types of exam.
	Retention	Learning with VP enhances retention.
	Mistakes	-Recognizing and correcting mistakes in a safe environment is crucial for successful learning. -VP mistakes are less prone to be repeated in clinical practice.
Teaching	Clinical specialties	-VPS should be used in all major clinical specialties. -Topics: frequent diseases and their complications; topics not included in the study plan and in the clinical rotations; diseases that might be easily missed during a short clinical rotation (due to seasonality or to being endemic in a different geographic area).
	Regulatory effect	-Institutional level: instruction becomes uniform across rotation sites. -Individual level: limiting the availability of the system externally regulates learning.
Assessment	Qualitatively different	-...and intrinsically better evaluation tool. -VPS should not be the only assessment form used in a course. Implicit: VPS should be used for course assessment only. -Allow increased retrieval of information in comparison with regular examinations.
	Motivation	VPS can increase motivation for learning.
	Professional focus	-Assessment and feedback on assessment are perceived as important learning tools. -VP assessment should be relevant for future clinical practice as a general practitioner.
	Production assessment	-Open questions make students think. -Students favour open questions even if the grades are lower.
Authenticity	Design and content	-Should reflect real clinical practice and offer localized menu / content choices. -Might consider including actual costs. -Artificial menus / content options are misleading.
	Localization of the socio-cultural context	Necessary for applications developed in one country and implemented in another.
	Realism and virtuality	-Real life records thought to make better patient cases than fictitious scenarios. -The knowledge derived from them is directly transferable to real patients. -Strong emphasis on patient photo. -Cases created from real life patient records no longer perceived as "virtual".
	Feedback	Actual patient evolution and effect of treatment are highly desirable features of feedback, adding to realism.
Implementation	Number of cases	More than one case per topic can be necessary for common diseases which are often complicated / have co-morbidities at presentation. Min. 5-6 cases of tropical diseases.
	Access and availability	The availability of the application should be restricted in time.
	VP exchange	Tropical diseases cases should be exchanged with other HEIs.

Table 4. Synopsis of themes and categories in study II.

In conclusion, studies I and II convey the importance of the following aspects: end-user customization; authenticity of the software design, clinical scenarios, media used to support the case and case feedback; use of VPS for clinical reasoning development, in a broad

curricular context of clinical specialties, supporting learning of topics not seen during clinical rotations; relevance of VPS assessment for future clinical practice.

Study III

The effect size ranged from 1.1 to 2.9. The mean score differences between the Web-SP and control groups in terms 1, 2 and 3 were statistically highly significant ($P < 0.001$). Both Web-SP and regular examination results were significantly higher in the study group compared to controls (P value < 0.001) in all terms. The paired mean difference in term 4 was also statistically significant for both haematology and cardiology ($P < 0.001$). A total of 216 students were enrolled in the four terms. The mean age ranged from 21.7 to 22.7 years. The median age was 22 years for both study and control groups in terms 1, 3 and 4, and 21 years in term 2.

The scores for the groups that used Web-SP were higher than for the control groups.

The unpaired mean score differences (mean score study group – mean score control group) and the 95% confidence interval for haematology and cardiology results by term are displayed in Table 5.

	Web-SP	Regular exams
Term 1		
Haematology	1.35 (0.85, 1.83)***	1.54 (0.86, 2.22)***
Cardiology	2.02 (1.57, 2.45)***	1.98 (1.42, 2.53)***
Term 2		
Haematology	1.29 (0.92, 1.66)***	2.61 (1.97, 3.27)***
Cardiology	2.15 (1.63, 2.68)***	2.89 (2.28, 3.50)***
Term 3		
Haematology	1.21 (0.86, 1.57)***	1.34(0.79, 1.88)***
Cardiology	1.07 (0.77, 1.36)***	0.90 (0.45, 1.37)***
Term 4 *		
Paired mean diff. †	0.66 (0.50, 0 .83)***	0.57 (0.45, 0 .69)***

Table 5. Mean difference and 95% confidence interval of assessment results of Web-SP and regular exams by term for haematology and cardiology.

* 95% CI (confidence interval) for the mean difference. † Paired mean difference and 95% confidence interval. *** $P < 0.001$.

The mean difference for Web-SP results (on a 0 to 6 scale) in term 1 was 1.35 (95% CI .85, 1.83; $P < 0.001$) for haematology and 2.02 (95% CI 1.57, 2.45; $P < 0.001$) for cardiology. For regular exam results, the mean score difference in term 1 for haematology was 1.54 (95% CI .86, 2.22; $P < 0.001$) and for cardiology 1.98 (95% CI 1.42, 2.53; $P < 0.001$). The mean score differences between the Web-SP and control groups in terms 2 and 3 were also statistically highly significant ($P < 0.001$). Both Web-SP and regular examination results were

significantly higher in the study group compared to the controls (P value < 0.001) in all terms. The paired mean difference in term 4 was 0.66 (95% CI .50, .83; P< 0.001) for haematology and 0.57 (95% CI .45, .69; P<0.001) for cardiology.

The mean scores of the study and control groups by term are presented in Table 6. The mean score for the study group ranged from 2.90 to 4.82 for haematology and 3.44 to 5.18 for cardiology compared to 1.10 to 3.61 and 0.73 to 4.11 for the control group for haematology and cardiology respectively. The maximum score was 6 points. The mean score was significantly greater in the study group for all terms (P< 0.001).

	Haematology			Cardiology		
	Study group mean (SD)	Control group mean (SD)	Effect size	Study group mean (SD)	Control group mean (SD)	Effect size
Term 1 (n=63)						
Web-SP	4.35(0.85)	3.00(1.07)	1.4	4.68(0.55)	2.66(1.10)	2.3
Regular exams	2.90(1.37)	1.36(1.32)	1.1	3.44(0.85)	1.46(1.29)	1.8
Term 2 (n=43)						
Web-SP	4.51(0.62)	3.22(0.59)	2.1	4.63(0.73)	2.48(0.96)	2.5
Regular exams	3.71(0.85)	1.10(1.12)	2.6	3.62(0.98)	0.73(0.99)	2.9
Term 3 (n=49)						
Web-SP	4.82(0.76)	3.61(0.43)	2.0	5.18(0.59)	4.11(0.41)	2.1
Regular exams	4.27(0.81)	2.93(1.07)	1.4	4.60(0.69)	3.70(0.90)	1.1
Term 4 (n=61) *						
Web-SP	3.61(1.03)	---*		3.77(0.89)	---*	
Regular exams	2.95(1.37)	---*		3.20(1.10)	---*	

Table 6. Mean scores, standard deviations and effect sizes for haematology and cardiology assessment results, by term. Maximum score is 6 points. *Term 4: no control group. It is a related sample of paired observation.

The mean (SD) score in cardiology for the study group with Web-SP assessment was 4.68 (0.55) compared to 2.66 (1.10) in controls in term 1. Similar results were observed for the

regular exam in the study and the control groups in all terms for both cardiology and haematology. However, the Web-SP results were higher in the study group, compared to the control group (p value < 0.001) in all terms. The overall Web-SP result of the study and the control group was higher than the regular exam result for both cardiology and haematology. The internal consistency, measured by Cronbach's alpha, ranged between .77 and .97. The intra-class correlation was .98 and Tukey's test of non-additivity was not significant ($p = .51$).

Study IV

The mean (SD) age was 22.2 (2.0) years for the study group and 22.3 (1.8) for the controls. Female participants accounted for 56% of the study group (14 students) and 62.5% (15) for the control group. The mean scores and standard deviations for exams performed on virtual patients (Web-SP) and regular exams are presented in Table 7.

Assessment	Haematology			
	Study group (n=25)		Control group (n=24)	
	Mean (SD)	Median	Mean (SD)	Median
Web-SP				
Early	8.2 (1.0)	8.1	6.8 (0.5)	6.8
Delayed	8.0 (0.8)	8.2	6.5 (0.4)	6.5
Regular exam				
Early	7.1 (1.3)	7.2	4.9 (1.8)	5.3
Delayed	6.2 (1.9)	6.5	4.3 (1.7)	4.4
	Cardiology			
Web-SP				
Early	8.7 (0.9)	8.5	7.3 (0.5)	7.4
Delayed	8.8 (0.9)	8.8	7.6 (0.5)	7.8
Regular exam				
Early	7.7 (1.1)	8.0	6.1 (1.5)	6.0
Delayed	7.9 (1.2)	8.0	6.1 (1.7)	6.5

Table 7. Mean scores*, standard deviations and median for the study and the control group by subject and assessment type. * Maximum score is 10 points.

The maximum score was 10 points. The mean score for the Web-SP and regular exam was greater in the study group for both early and delayed assessments for both topics. The effect size ranged from 0.5 to 0.8.

The mean score differences between Web-SP and control groups and the 95% CI for haematology and cardiology results by assessment time are displayed in Table 8.

Assessment	<u>Haematology</u>	<u>Cardiology</u>
	Mean difference (95% CI) ^a	Mean difference (95 % CI) ^a
Early (n=49)		
Web-SP (n=25)	1.43 (0.96, 1.91) ***	1.34 (0.93, 1.76) ***
Regular exams (n=24)	2.21 (1.3, 3.13) ***	1.52 (0.76, 2.28) ***
Delayed (n=49)		
Web-SP (n=25)	1.48 (1.09, 1.86) ***	1.16 (0.74, 1.58) ***
Regular exams (n=24)	1.96 (0.93, 2.98) ***	1.74 (0.89, 2.58) ***

Group	Mean difference (95% CI) ^b	Mean difference (95 % CI) ^b
Study group (n=25)		
Web-SP	-0.18 (-0.42, 0.06)	0.10 (-0.13, 0.32)
Regular exams	-0.89 (-1.40, -0.38) **	0.18 (-0.13, 0.48)
Control (n=24)		
Web-SP	-0.23 (-0.39, -0.07) **	0.28 (0.09, 0.46) **
Regular exams	-0.63 (-1.60, 0.34)	-0.04 (-0.92, 0.84)

Table 8. Mean difference and 95% confidence interval for haematology and cardiology of early and delayed assessment results. ** P<0.001, *** P<0.001.

^a 95% CI (confidence interval) for unpaired mean difference (study mean score - control mean score).

^b 95% CI for paired mean difference (delayed mean score - early mean score).

Statistically highly significant between-group and within-group differences were observed both for haematology and for cardiology. The mean difference for early assessment Web-SP results (study mean score – control mean score) was 1.43 (95% CI .96, 1.91; P< 0.001) for haematology and 1.34 (95% CI 0.93, 1.76; P< 0.001) for cardiology.

For regular exam results, the mean score difference for early assessment was 2.21 (95% CI 1.30, 3.13; P<0.001) for haematology and 1.52 (95% CI 0.76, 2.28; P<0.001) for cardiology. The difference in mean score for delayed assessment Web-SP results was 1.48 (95% CI 1.09, 1.86; P< 0.001) for haematology and 1.16 (95% CI 0.74, 1.58; P< 0.001) for cardiology. For regular exam results the mean score difference for delayed assessment was 1.96 (95% CI 0.93, 2.98; P< 0.001) for haematology and 1.74 (95% CI 0.89, 2.58; P<0.001) for cardiology.

The within-group mean differences (delayed mean score – early mean score) for the study and control groups are displayed in Table 3. While scores or differences in mean scores for delayed retention were lower than for early retention for both groups, the study group's performance remained better than the control at a statistically significant level. Statistically

highly significant paired mean differences were observed for the control group in Web-SP results in both haematology and cardiology ($P < 0.001$).

The mean ranks of study and control groups by patient interview, laboratory, diagnosis and treatment categories for Web-SP assessment are presented in Figure 8, illustrating that the study group scored higher in all categories. Significant between-group differences were observed in all the categories ($P < 0.01$), except for patient interview in early assessment ($P = 0.08$) and laboratory in delayed assessment ($P = 0.08$), in haematology.

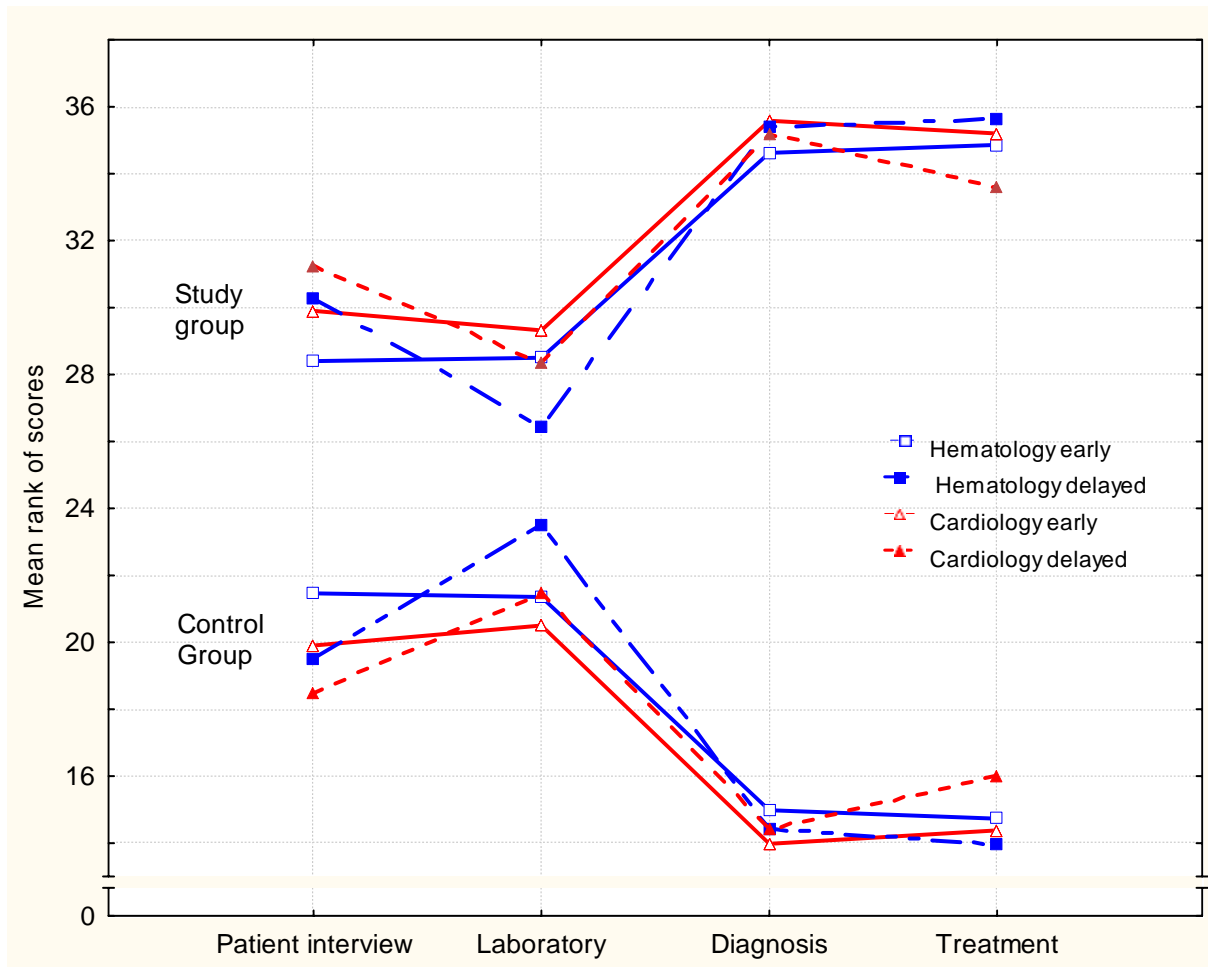


Figure 8. Line plot of mean rank of study and control groups by patient interview, laboratory, diagnosis and treatment categories for Web-SP assessment.

In a nutshell, the main findings of studies III and IV were that the evaluation results for the Web-SP group were superior to the control group, both in virtual and regular exams, and that delayed retention was higher in the Web-SP group, compared with the control group. Assessment with VPS seems to yield overall better educational results than regular course examination, when teaching is constructively aligned with evaluation.

DISCUSSION

Studies I - II

Study I dealt with the opinions of the stakeholders - medical students, clinical teachers and faculty board - on the implementation and use of a VPS. Interestingly, no systematic differences were observed between the opinions of stakeholders, regardless of country; despite the notable differences in educational setting, language and case origin, the participants, faced with the same problems, arrived at similar solutions and conclusions. Study II explored the perceptions of medical students on the use of a VPS, by means of a qualitative approach - focus group interviews. From the array of qualitative methods, both questionnaires and individual in-depth interviews would have been reasonable alternatives. As we intended to collect as much good-quality data as possible while still keeping the project feasible, we chose the focus group interviews. The students also came up with novel information, i.e. explorations and descriptions that brought up issues previously unknown to the researchers, which might not have surfaced with individual interviews or questionnaires.

The main findings of these studies are presented in the following pages as elements of the Issenberg product (*training resources x trained educators x curricular integration = effective simulation training*) [36].

I. Training resources - the VPS

The VPS interface and the virtual cases

According to previous research, the *design* of the VPS applications should reflect the reality of clinical practice [34, 5] and offer localized menus and content choices [28].

Customization and sustainable development

The ease of customization by end-users seems to be critical for successful VPS implementation. All stakeholders directly involved in case creation expressed the need to extensively edit the Web-SP “shell”, in order to adapt it to specific learning objectives or to a certain sub-specialty. This would require removing features considered unattractive or unnecessary (e.g. parts of the physical exam section) and adding new features, or grouping the existing ones in different ways (e.g. labs and imaging), to support the learning objectives of the actual course. Completing a case in accordance with the provided matrix was regarded by staff with heavy clinical burdens as time-consuming and ineffective. The stakeholders

consider that VPS should either allow extensive editing by end-users or be subject to post-implementation development by the institution of origin; the latter is in keeping with opinions expressed in the literature [30].

Authenticity

The importance of authentic virtual patient contexts in terms of interface and user tasks was flagged in a recent article on the CAMPUS system [34]. The interface of the physical examination section should “help students proceed in a structured manner”; the system should also allow students to make similar decisions to those “a real doctor would make”. The participants in our studies emphasized as well the connection between software design, case content and authenticity. A physical exam section that puzzles both case creators and students should be a target for improvement. All aspects of design should support clinical reasoning; in our case, the stakeholders exemplified the lab and imaging section and suggested those be clustered in groups of tests. However, since such “batteries” differ from country to country, the adjustments should be subject to localization efforts.

Nevertheless, there’s more to VPS authenticity than meets the eye. A need for localization of “adopted” VPS to the medical practice and the socio-cultural context of the host country has been advocated in the literature [28]. The case scenarios and the nature and quality of feedback can add to or subtract from the authenticity of an application. The students feel that the knowledge derived from cases based on real-life patient records is immediately applicable to the clinical setting, contributing to enhanced motivation in using the system. Another feature highlighted by students is the quality of the feedback, as the Spanish version of Web-SP offers the real treatment course and the actual clinical evolution of each patient. For an adopted application, native language may add to the perceived realism. The Spanish-speaking students in our study indicate they would not like to use an application in English; some even object to patient cases reflecting medical practices different from the context in which they are supposed to work as physicians in the future.

Layouts that feel artificial in terms of menus or content do not meet their educational goal and are considered by students as downright misleading [5]. In contrast, knowledge derived from virtual cases created from real life clinical records is thought to be *directly transferred* to actual patients, which is a novel finding. An additional possible benefit of using real cases to create virtual patients is a further enhancement of authenticity to such a level as to consider the application as a surrogate for reality (a desirable feature in the case of rare diseases, topics

not seen during a clinical rotation, diseases unavailable geographically or seasonally, etc.). Feedback derived from real cases greatly adds to the realism of the application.

The literature [19, 34] supports the assumption that authenticity can be conveyed by means as simple as the face photo, together with shots of the main findings of the physical exam. The students' emphasis on the importance of the face photo is a novel finding; nevertheless, the Spanish version of Web-SP features photos of genuinely ill individuals, like those who may well wait in line for a consultation at the clinic. The students respond empathetically to faces they consider familiar. More media resources are actually rarely needed. According to the experience at Karolinska Institutet, the creation of such cases is unnecessarily expensive and their added value is minimal. However, informed consent for any media use is an obvious must for cases created from actual clinical records, as well as for the content of the medical history itself.

VPS use in teaching and assessment

VPS for teaching and learning

There is a broad literature consensus in regarding "clinical reasoning" as the best use of VPS [19, 34, 28] and our findings support it. The students, however, differ from leaders and teachers in assuming that VPS should offer a reasonable load of factual knowledge with each patient ("factual and core knowledge" input via the virtual patient application is considered by students as a highly desirable feature of VPS).

More surprising is the students' preference for usual presentations of common, frequent diseases, particularly cases already complicated at debut, patients with several co-morbidities, as well as cases of drug interaction, by contrast with another study which had previously reported a student inclination for problems they were "unlikely to encounter during clinical training" [34]. The recent accreditation requirements of the Liaison Committee on Medical Education perceive VPS as appropriate for teaching topics not seen during clinical rotations, due e.g. to seasonality [25, 59]. However, it is interesting that more cases equate to more knowledge in the students' opinion.

Again surprisingly, the students evoke a regulatory effect of VPS at individual level (to help plan their learning) and at institutional level (to even out the differences in disease range among rotation sites).

Instead of re-experiencing the daily frustration of not knowing what happened with their patients (a rotation ended, the patient was moved to another floor, etc.), the students convey the importance of getting closure by means of patient feedback.

The students indicate that the *transfer of knowledge* to the real patient is the ultimate goal of simulation technology, in agreement with the literature [41, 34, 6].

Medical students feel they remember more with VPS. *Retention enhancement* with VPS has previously been demonstrated in the same cohort of students, where the effect of VPS on early and late retention was studied [6].

As for mistakes, free from a stressful context (patient, family, hospital staff), the students still perceive errors as serious events, but at the same time as meaningful learning opportunities, to the point that they consider their repetition as unlikely in their future clinical practice.

Assessment

As for the assessment component, the students consider VPS a good assessment tool [7, 6] and the stakeholders envisage using VPS as a course examination, in combination with other assessment methods. To come across the students' conviction that VPS assessment is qualitatively different in comparison with other evaluation methods is a novel finding.

A further benefit of a virtual patient application is that assessment may not feel like an exam, leading to an increased motivation for learning itself. The students are also aware that they can and should learn from assessment; feedback is deemed crucial for such learning through assessment to occur.

The students do not support, however, a VPS assessment directed to anything other than the knowledge and skills essential for clinical practice as a general practitioner, or the use of VPS outside regular course evaluation. An immediate use of VPS for high-stakes examinations is not obvious to the stakeholders, a finding we link to two Web-SP features: 1) the absence of an automated score feature, to ensure reliable and reproducible assessment results; 2) student-identity blinding is currently impossible in Web-SP. Caution should therefore be exercised when generalizing a benefit for learning across different systems, as the assessment formats are different (open-ended questions in Web-SP, multiple choice in other systems); feedback, if provided, largely varies in layout.

In a position conflicting with the scope of a web-based application – namely round-the-clock access from any site - most students recommend that the availability of the application be restricted to certain term timeslots, as a means of external regulation. This might reflect special characteristics of the student population, whose identification was beyond the scope and the methodology of the respective study.

The exchange of virtual patient cases is not a priority for many institutions [5] and time will tell to what extent current inter-operability efforts are worth undertaking (e.g. reports of number of user sessions per case exchanged are awaited).

Educational setting

The students in our study had used VPS both for learning and for assessment in their Internal Medicine course, where Web-SP was a curricular component. Not surprisingly, the Colombian stakeholders (students included) do not envisage a use for VPS outside of the curricular context. The students recommend that other major clinical specialties offer the opportunity to work with VPS; especially so if, as mentioned earlier, certain topics are not included in the study plan, but are likely to surface in the first level of patient care.

The stakeholders also agree the learning objectives must differ according to the context (different programs at undergraduate level or undergraduate vs. post-graduate setting); the participants are consistent in finding that the level of difficulty should vary, as well as the type and depth of feedback provided; even the flow through the case should be adapted to the desired competence level (the more sophisticated the formation – continuing medical education, resident training - , the less appealing the linear design).

Post-implementation activities

Not surprisingly, the stakeholders do not agree when asked about post-implementation activities that best support the use of VPS. They are united, though, in considering that the continuous creation of new cases is essential to the post-implementation setting and at least the teachers believe that peer validation is equally important to new case creation. Some degree of consensus on the necessity of having a functional system (mainly as Internet connection, logins, IT support) exists among stakeholders.

The leaders awarded low scores to the item “case use in a curricular context” (the highest ranking item for teachers). Leaders also gave very low scores to both case validation and case

exchange with other higher education institutions. The leaders' lack of interest in case peer review and validation may reflect a *de facto* conviction that it is the VPS "shell" that validates the case content. Without a proper case-script, aligned with educational objectives and with assessment, case creation limited to filling data into empty fields, performed by staff other than clinicians, poses a serious threat to content validity.

Costs

Intimately connected to the continuum of implementation, VPS development *and* maintenance costs are crucial to the successful use of an application. Production costs per case vary largely in a review on VPS use in North America [33]; more than half ranged from 10,000 to 50,000 USD. Another review article found that 85% of cases cost more than 10,000 USD per case [19]. In the CLIPP project, the estimated overall development cost per case was 18,000 USD [26]; this is the only paper to mention maintenance costs, in the region of 120,000 USD annually. Even if we do not agree that the cost for developing a single case is as high as these figures [62], costs can be a problem. In the present studies, the costs generated by the implementation process and by the maintenance of the application rank low on everybody's agenda; however, the leaders seem more cost-conscious than other stakeholders.

II. Trained educators and staff development

Who will develop and maintain virtual patient cases is no trivial matter [19], bearing on VPS overall costs. As mentioned earlier, effective training in simulation was postulated in 2006 to be "the product of training resources, trained educators and curricular institutionalization" [36]; nevertheless, staff development in the field of simulation has hardly been discussed in the literature [56, 30, 46]. The need for resident training in simulation-based education was recently identified [46]. Staff development to support VPS use was regarded by leaders as important to the post-implementation setting, but it was not mentioned by any other stakeholder category.

III. Curricular integration

Curricular institutionalization of simulation is the third component of the Issenberg product. Should those VPS applications not integrated in the curriculum be considered *a priori* as non-effective? The systems implemented as add-ons to the existing curriculum have few users indeed. In their article on the experience with the CAMPUS system, the authors acknowledge that "the usage...has not fulfilled the expectations" and that "as soon as students get credit

points for completing case sessions most of them will do it” [30]. In opposition, the CLIPP project, where the CASUS system was used to teach the North American core paediatrics clerkship curriculum, reported more than 8,000 users in 98,000 sessions [26]. The present research envisaged both sides of the coin, as the same VPS was part of the curriculum in Colombia and an add-on in Sweden (with compulsory and voluntary use, respectively). However, regardless of Web-SP setting, curricular integration is still perceived by all stakeholder groups as crucial to reaching the educational goals set for a course.

Studies III - IV

As mentioned before, studies I and II also brought an insight on assessment issues. In the stakeholders’ opinion, VPS assessment is congruent with ordinary course evaluation in combination with other examination tools, and should not be used for high-stakes exams. Medical students, in particular, see VPS assessment as qualitatively different from regular exams and appreciate the fact that it is a production assessment, focused on essential aspects of their future profession. As all studies progressed, it became clearer and clearer that VPS use for assessment is circumscribed to the more encompassing concept of the *continuum of implementation* (the controlled, sustained implementation and educational use of the system, during the entirety of its useful life)⁶.

As for study III, it yielded both expected and unexpected results. The study group performed better in Web-SP exams than in regular exams, both in haematology and cardiology, with a trend for higher marks in cardiology. The control group *also* performed better in Web-SP exams than in regular exams. It is of concern that students may be prompted towards the correct diagnosis by the many features displayed by the VPS, especially in the laboratory section. Web-SP has a linear design and actually allows students to go back to previous sections; they can also order as many diagnostic tests as they deem necessary. Such features can introduce a significant variation from the paper-case, where no additional information can be obtained. However, there is evidence in the literature that the extent of laboratory work-up has no correlation with the degree of diagnostic accuracy [27].

The main finding is that the study group performed better than controls both in Web-SP and in regular exams, which advocates for the superiority of the virtual patient application. The blended approach in term 4 gave lower results than terms 1, 2 and 3, both in haematology and in cardiology. We attribute the inferior results observed in term 4 to its paired design,

⁶ The concept is further developed in the *general discussion*.

resulting in lower mean differences than those of the parallel design in the other cohorts. Additionally, the change of the course director might have meant a different approach to conveying the importance of using Web-SP in the Internal Medicine curriculum and therefore decreased interest of the students in the new educational tool. This suggests VPS may be context-sensitive.

The controls in term 2 showed an anomalous behaviour in the regular exams, with extremely low marks in the paper examination; unfortunately we could not explore its origins during data collection, other than by informally discussing it with students. Those in the control group manifested a certain resistance to evaluations made in “the old way” and believed Web-SP to be a better assessment tool, while the study group felt they were better equipped to deal with paper-and-pen exams after the virtual patient encounters. This may indicate that when offered novel assessment modalities, students might actively oppose other types of examination they consider unattractive; the format of the assessment instrument and the student motivation may both play a role in such settings.

At the time the studies were performed, Web-SP did not feature an automated-scoring system (in the current version, a preliminary assessment module is available). Scoring was therefore manual and based on the specific scale developed for this study. Similar experiences are registered in literature [53, 27], while automated models are so far envisaged for summative assessment [20, 61]. A major downside of manual scoring in larger cohorts of students is that it is time-consuming and therefore difficult to replicate by another researcher or faculty staff.

It may be concluded that VPS use both for learning and for assessment in the same course is a good way of supporting student learning; this is the setting where Web-SP assessment results were consistently superior to those obtained with regular course assessment. However, the delivery of the same learning tool (VPS) might prove to be context-sensitive and care should be exercised when generalizing results. Moreover, learning with VPS could lead to better assessment results with different examination formats and ultimately provide a transferable skill. We believe that the transfer of the knowledge and skills to live patients is the ultimate goal of simulation-based medical education, and that it could be achieved not just by procedural simulation [48], but also with VPS learning and assessment.

In study IV, the study group showed similar results between the early and delayed assessment with virtual patients (Web-SP arm) for the haematology topics, suggesting a good retention of knowledge at four months, despite the absence of formal clinical rotations in haematology. In

regular exams, the study group demonstrated a moderate loss of knowledge in the delayed assessment; however, their grades were still superior to the controls'. The control group obtained lower grades overall than the study group, with significant loss in delayed examination with traditional methods.

The cardiology results were similar. The study group had a small gain in delayed examination results with Web-SP and no or small gain in regular exams; nevertheless, they scored much better than in haematology topics examined with traditional methods. Again, the grades obtained by the study group were superior to other results of assessment, regardless of group assignment. The controls demonstrated knowledge gain in both modalities of delayed assessment, with a significant superiority of Web-SP, which may be related to solid clinical rotations in cardiology between the first and second round of examinations. Even though the controls lacked the learning component with virtual patients, they scored better in the Web-SP examination, compared to the regular course assessment.

VPS can enhance learning efficiency in several ways. In this second study, the early assessment was conducted after an initial period of massed distribution of learning, which is known to yield good test results [40]. In the "bolus education" period the study group used virtual patient cases for learning and obtained higher results in the early assessment, irrespective of the testing modality. The use of interactive computerized modules is associated with improved retention in the clinical setting [9]. Moreover, if presented concomitantly with traditional learning methods, the results of tests are at least as high as with standard education alone. In our study, the participants were randomized either to VPS learning or to traditional learning, in order to minimize the confounders.

Feedback, one major advantage offered by VPS, boosts self-assessment [10], which allows repeated retrieval of knowledge, re-modelling and re-encoding of knowledge. In our study, the feedback module was accessed immediately by the students, but a variable delay can generally be associated with superior retention [44].

Knowledge decline is in the range of one-fourth to one-third after a retention interval of 12 months. Its rate is not related to the speed of acquiring the information, but to the type of practice after learning [38, 50]. Knowledge loss does not become a problem if frequently retrieved in a repetitive manner, through testing, practice (such as clinical rotations) or both. In the cardiology setting the participants benefited from structured rotations; as such, both groups performed better in late assessment (with virtual patients and in regular examinations), a finding consistent with the role of the clinical rotation in maintaining and increasing the

knowledge in the subject area. The opposite trend was observed for haematology topics, with important loss in delayed assessment for the control group. The use of Web-SP for learning and assessment in the study group seems to have had a positive influence on retention in the absence of structured clinical rotations. Similarly to the literature [39], the control group displayed poorer retention in the absence of Web-based teaching, even in students who underwent a structured, domain-specific, clinical rotation.

In agreement with the literature [40], these findings suggest that VPS should be used for spaced education (as opposed to massed distribution), followed by assessment rounds; as such, VPS will lead to efficient learning with improved knowledge retention. The so called “bolus education” improves early performance tests, while spaced distribution and frequent testing would lead to better results on delayed assessments, with superior effects on long-term retention of knowledge [50, 38].

To sum up, the findings indicate a better delayed retention with virtual patients than with traditional learning methods, despite the limited number of virtual cases available for learning in the study group. In the approved course curriculum, the learning component was condensed in the beginning of the term. However, spaced learning with frequent assessments would have perhaps yielded an even better retrieval of knowledge. Nevertheless, this approach to learning and assessment also has a “downside”: it demands a large pool of clinical cases, either of own production or from repositories. For cases drawn from repositories, localization could become a necessary step, in order to maintain the realism of the application and the content validity. Regardless of the source, case quality is paramount here and to our knowledge has not been the object of special interest in the literature. The robustness of the software together with effective case design, appropriate flow through the case, necessary alignment of the learning outcomes with the assessment modalities, right timing for feedback, and finally, the essential curricular integration with a proper temporal sequence within a given course are all decisive for the results of any assessment with VPS. Higher education institutions that have implemented VPS should use them regularly in student course evaluation.

A more general discussion of the implementation and use of VPS systems

Let us re-examine the Issenberg postulate on effective simulation training. VPS undoubtedly *are* valuable **training resources**, and the studies on assessment results included in this thesis are a proof of their efficacy. However, the educational effectiveness of VPS might be context-sensitive and require a controlled *continuum of implementation*. The VPS software design and the virtual cases themselves should comply with certain requirements, in order for effective learning to occur.

VPS design. From the early creation phase, VPS design should be adapted to the proposed learning objectives and to the end-users' needs, and not vice versa. The design should seamlessly foster clinical reasoning development and not try to appeal to skills which are better developed by other means (e.g. history-taking skills or communication skills). The design should not challenge common sense (e.g. auscultation of the elbow from the front) or the established medical practice of the host institution / country. The design package must deliver a sense of realism, while still keeping design options simple.

When designing VPS, it can be a good policy to remember the targeted end-users and their needs. Bear in mind that student motivation in using VPS might be grossly overrated. Today's students, who spend hours on YouTube and Facebook, might prove themselves unwilling to spend 15 minutes on a badly designed, and otherwise boring, VPS interface. Inflexible design choices and the lack of on-demand customization will certainly drive clinicians away as well. Moreover, if there is no customization margin, the targeted educational settings cannot be numerous.

Any design should today include two features: feedback and automated scoring. Feedback surely adds to authenticity, but also fosters student learning, while automated scoring might boost the use of the VPS in assessment (and thus effortlessly align learning with assessment).

The virtual cases. The creation and the educational use of virtual patient cases raise a few quality issues. The creation should be aligned with the proposed learning objectives; the cases should be peer-reviewed; the virtual patients should be used in a curricular context, both for learning and for assessment; old(er) cases should be updated, to remain in line with the current standards of practice; new cases should be created at a constant pace, so that a pool of cases, ensuring sufficient clinical variation, is finally at hand. The HEIs should be aware that certain categories of learners may not benefit from cases created in other healthcare contexts

or from cases presented in a language the users do not master; moreover, some learners may require cases of different design and content than their fellow learners from similar programs / institutions / countries.

Case creation is a combination of narrative availability and clinical experience of the creator. A choice adding to the realism and authenticity of the VPS applications is the collection of cases from real clinical records. The advantage is the obvious credibility and accuracy of the clinical scenario, offering the end-user the experience he needs in real life (an opportunity greatly appreciated by students); the downside is the tremendous amount of work needed to collect such cases and their related media, not to mention the imperative of both patient informed consent and approval of the healthcare institution.

However, a combination of credible clinical scenarios, media and feedback, supported by excellent interface design, becomes a recipe for success. According to undergraduate medical students, such cases may well develop clinical reasoning paths applicable to real patients.

VPS use for assessment and the life after Biggs. More than a decade after the birth of constructive alignment, the concept became an almost axiomatic truth in higher education. When teaching, assessment and the intended learning goals are aligned with each other, the educational results are better in term of grades and, more importantly, conducive to deep learning. The students do tend to structure their learning activities around assessment, in an effort to optimise their performance in exams. VPS are no exception: clinical reasoning development and better assessment results are unlikely to be obtained by simply adding a few cases to an existing course, otherwise lacking constructive alignment. VPS should be implemented both for teaching and for assessment and the learning outcomes should constantly be adapted to the emergent reality of the classroom.

Even in a constructively aligned curriculum and under rigorous conditions - localized application, embedded in the course plan, delivering peer-reviewed, realistic cases with rich feedback -, a trivial event, such as the change of the course director, may trigger inferior evaluation results, as seen here with term 4 in study III. Such findings put into perspective the issue of teacher motivation in using information technology, but also the “stability” of the system in an otherwise very controlled instructional setting; we should bear in mind that the level of entropy of the “normal” classroom teaching / evaluation environment might be higher than that of a research project. Superior educational results with VPS may indeed require the careful, timely and sustained adjustment of many variables over a long period of time, i.e. conditions seen only in research projects. However, when all the stars do align, the end result

is worth the effort: the students learn better, do better regardless of examination type and seem not to forget as much.

Localization and customization of VPS. The *localization* of an adopted application is almost imperative. Otherwise the end-users might question its learning benefits, consider that the application lacks realism and even refuse to use it. Who should be responsible for localization (the institution of origin or the host) remains an open question.

Customization - adaptation to end-user requirements - is a sensitive topic as well. A VPS too generous in its design, created to appeal to too many user types, may end up accommodating very few. To give an example, an obstetrical case may well need a different layout than an infectious disease or an Alzheimer's case. Easy end-user customization must be allowed for already in the design phase of the future application. Such an approach transfers the burden of customization onto the user, not the developing team / mother institution.

So the successful implementation and use of a VPS application requires a number of different, complex, and concerted interventions in the implementation - post-implementation setting (the continuum of implementation). These actions, all essential for the survival and the proper educational use of any application, are governed by people who may have different perceptions on the usefulness and the ease of use of a particular IT system. All **stakeholders**, and especially the end-users, play a crucial role in determining the fate of a VPS application, as they may embrace it and use it consistently, or reject it. Underuse, resistance, work-around the system or the newly acquired roles and tasks, sabotage and even abandonment are according to biomedical informatics journals encompassing problems of healthcare information technology; anecdotal evidence suggests this might also be the case with VPS applications.

Given that so many people are involved in so many ways in the continuum of VPS implementation, the second element of the Issenberg product, the **trained educators**, should probably be replaced by the more generous concept of **stakeholders**. The stakeholder group, far from being homogenous, could be divided into: *i*) developing team (which should include trained educators); *ii*) direct users: clinical teachers on one side (trained educators); undergraduate students, residents or physicians enrolled in CME programs, on the other side; *iii*) indirect users: origin / host institution, academic community, medical community. The VPS community should be aware that *i*) most stakeholder categories are directly accountable

for VPS implementation and use; *ii*) **all** stakeholder groups play a part in the acceptance of the application.

IT acceptance by stakeholders does not lack a theoretical frame. The so-called *Technology Acceptance Model* (TAM) can also be applied to healthcare IT, in order to predict and explain the end-users' reaction to system implementation and use. Based on Fishbein and Azjen's *Theory of Reasoned Action* - a very general model that in 1980 suggested beliefs influence attitudes, which determine intentions, which in turn dictate behaviour - , TAM was proposed by Davis to predict system acceptance and to diagnose design problems (Fig. 9).

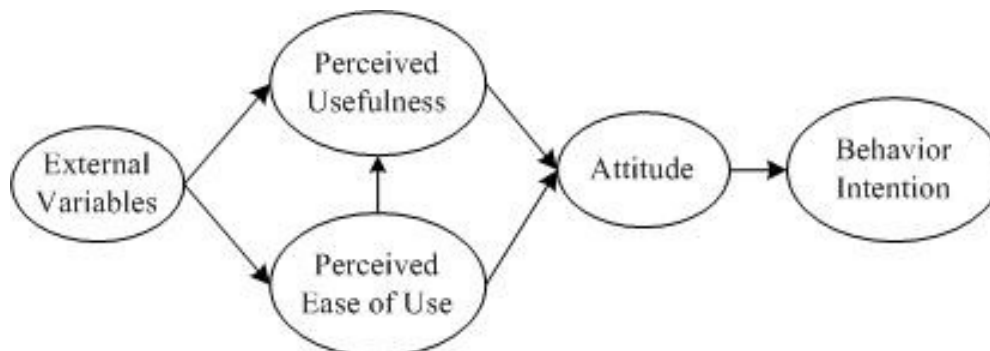


Figure 9. The Technology Acceptance Model (Davis FD, 1989: "Perceived usefulness, perceived ease of use, and user acceptance of information technology". MIS Q. 13:319-40).

According to TAM, the intention to accept and use technology is determined directly by attitude, perceived usefulness and perceived ease of use. Perceived usefulness refers to an individual's perception that the utilization of a particular technology will be advantageous in an organizational setting over the current practice. Perceived ease of use refers to the assumption that the utilization of the new technology will be effortless. As shown by literature and as conveyed by the results of the present research as well, both components crystallize into individual attitudes towards the new system, which culminate with the intention to use the technology, and with its actual use or disuse.

A short comment on the necessary duality case creator – clinical expert. Who the storyteller is may well determine the quality and educational value of a virtual case. An experienced clinical teacher, familiar with the curricular requirements, in possession of a credible narrative and of adequate media to back it up, is capable of aligning the storyline with the learning objectives and with the characteristics / learning needs of the student population he is knowledgeable of. Higher Education Institutions might want to adopt VPS for the wrong reasons, e.g. to reduce staff costs, and may require other categories of staff (i.e. other than

clinicians) to “tick boxes” in a virtual patient frame. This is a truly unfortunate choice for staff development, student learning and ultimately for the fate of the application. Considering the post-implementation continuum, costs are unlikely to go down after VPS adoption, especially given the low life expectancy of the application. HEIs should let experienced clinicians create good quality cases on decent software; such cases might augment the lifespan of the application and give reimbursement a fighting chance.

Curricular integration is something everyone agrees upon when talking about VPS.

Unfortunately, it is easier said than done. Let us not forget that a VPS running as a pilot or as an add-on is not educationally effective and adds to the current confusion on virtual patient efficacy as a learning tool. So no matter how hard curricular integration might prove to be, VPS still should not be offered outside a curriculum.

The faculty boards should be aware that VPS cannot entirely replace other learning tools. Their use for learning should support clinical reasoning development across clinical specialties, presenting either the core curriculum or topics not delivered by the curriculum (such as diseases not seen during a rotation due to seasonality or to geographical unavailability of patients). Curricular flexibility is certainly advisable and should allow for specific learner needs - such as common diseases in usual presentations, the case of the student population in this research.

The same VPS application, supported by a built-in scoring tool, should be used for regular course assessment, thus aligning learning with assessment. Both learning and assessment with VPS should be spaced and allow for multiple opportunities for information retrieval.

Finally, *the components of the Issenberg product do not exist in a cosmic void*. Neither the VPS application, nor the stakeholders, can be separated from their instructional context or from the local medical and socio-cultural practices, which are intertwined in a complex milieu. This nutritive or poisonous environment becomes an essential component of the continuum of implementation (along with e.g. VPS further development, authenticity, localization and customization; continuous creation and maintenance a pool of good quality virtual cases, meeting the learning objectives; curricular integration; educational results; accountability and sustainability).

The Issenberg product applied to virtual patient simulation would then look like this:

VPS as a training resource x stakeholders x continuum of implementation =

effective learning with VPS

CONCLUSIONS

Authenticity of virtual patient design and clinical scenarios

The need for a VPS design enhancing clinical reasoning abilities and supporting case authenticity cannot be overemphasized. Authenticity, however, extends well beyond the design of the interface. The users are more positive and also benefit more from an application whose case content is robust, derived from everyday practice and supported by feedback providing an exposé of actual patient treatment and evolution. For it is in this particular context that medical students perceive virtual patients as learning and assessment tools fostering clinical reasoning, able to facilitate the transfer of knowledge to the real patient.

Sustainable development

VPS applications should be allowed to evolve during their useful life and should be discontinued when obsolete. End-user customization, the localization of adopted applications, the consistent use of VPS for constructively aligned learning and assessment in a broad curricular context of clinical specialties, together with the creation of a pool of clinical cases relevant for the future clinical practice and meeting the educational needs of the target population, are the pillars of the *VPS continuum of controlled implementation*⁷. The use of VPS for assessment is a necessary consequence of teaching / learning processes with the educational tool, leading to superior evaluation results irrespective of exam type and to better delayed retention. These are indeed burning practical issues, circumscribed to the sustainable development of the field of virtual patient simulation.

Accountability

Accountability is crucial for sustainable development as well. VPS creators and users should report to the academic community both their joys and their sorrows with these systems. They should inform on the less glamorous topics, such as arrested development, lack of curricular integration, insufficient number of users, short useful lifetime of an application, high direct and indirect costs of development and maintenance, results of localization and customization efforts, continuous staff development and possible implementation errors. They should also

⁷ The continuum of implementation could be defined as *the controlled, sustained implementation and educational use of the system, during the entirety of its useful life, as well as the intricate pattern of interactions established during the lifespan of a VPS application between the stakeholders and the new technology, ending in its use or misuse.*

provide an accurate account of the life and times of a VPS - once the application is no longer in use - so that other educators and developers may benefit from their expertise.

The future is (still) bright

However, developers, educators and researchers will soon have to deal with issues such as VPS authenticity and end-user customization, if VPS are to *stay* in the mainstream of medical education. For the same reason, accountability and sustainable development profile themselves as imperatives for the virtual patient simulation field.

FINAL WORDS

My students were both the reason and the means for my research. These last lines go to my fellow physicians, who teach medical students and use virtual patients in their clinical course.

Students expect a virtual patient simulation system whose design and content is explicitly directed towards the development of clinical reasoning. They cannot quite put their finger on it, but would like to see it used both for learning and for assessment, in an overall important clinical course, such as Internal Medicine. Or, better still, in all major clinical courses. They learn from applications that are honest and clean-cut in their design choices and media support. It might be said that in terms of design, less is more; in terms of content and feedback, it is their richness, inner structure and realism that ensure the painless transfer of the knowledge acquired virtually today to the patient waiting in line tomorrow. Students also expect to face *a patient*: a living, breathing, sick person, not a software concoction starting with a CV photo for the late Lehman Brothers. Then, and only then, will they remember the disease behind the patient.

These are indeed great expectations. If we cannot fulfil them by means of virtual patients, we should orient our clinical teaching towards other - old or new - horizons.

ACKNOWLEDGEMENTS

At the end of the road, I look back and remember those who travelled with me. It is with deep gratitude and affection I give thanks to:

- Uno Fors, who was blissfully unaware of what he was getting into when he answered an email in 2004. A research project started soon after at my home university and the journey began. Later on, when I became a doctoral candidate at Karolinska Institutet, I was privileged enough to have Uno by my side as my main supervisor.
- Håkan Hult, who joined this project in 2007 as my co-supervisor. His expertise and his gentle manner were always greatly appreciated. It is only fair to say that without his advice and guidance I would never have considered qualitative methodology when designing the studies on implementation.
- Mesfin Kassaye Tessma, competent statistician and reliable co-author, who also happens to be a very patient and wise man. My research was genuinely improved by his input.
- Alban Office in Portugal and their EU Scholarship Programme, whose financial support made the first three years of my doctoral studies possible.
- Läkemedelsverket, my employer, for having a sound policy on professional development. To Ingrid Wallenbeck, head of the Clinical Trials Department, and to my fellow colleagues there, for ensuring proper work conditions to finish my research.
- Universidad el Bosque, for making all of this possible. To colleagues and dear friends I left behind there. Among them, Dr. Hernando Matiz Camacho, my external mentor, ex-Rector and Dean, for it is a rare privilege to count your former boss among the restricted circle of personal friends. And, of course, to my students.
- The total strangers who came into my life in this new country and helped me when I most needed help: to Lars and Ingela Andersson, Suzanna and Cornel Vendel, Italo Masiello.

- Those who always had a smile and a kind word for me at LIME, Karolinska Institutet.
- My light into the darkness, my family: for their faith in me and for supporting my decision to change their lives. May we travel together many more roads.

At the end of the journey, I remember as well those who tried hard to make this an impossible mission. Those people are usually not given the proper credit in one's accomplishments. Thus, in the spirit of fairness, I would like to take this unique opportunity and acknowledge their decisive contribution to reaching my goal, as they were indeed creative and resourceful: thank you for showing me the road.

REFERENCES

1. Barber SJ, Rajaram S, Marsh EJ (2008). "Fact learning: how information accuracy, delay, and repeated testing change retention and retrieval experience". *Memory* 16 (8):934-46.
2. Bergin R, Youngblood P, Ayers MK, Boberg J, Bolander K, Courteille O, Dev P, Hindbeck H, Leonard II EE, Stringer R, Thalme A, Fors U (2003). "Interactive simulated patient: experiences with collaborative e-learning in medicine". *Journal of Educational Computing Research* 29 (3):387 – 400.
3. Bergin RA, Fors U (2003). "Interactive simulated patient – an advanced tool for student activated learning in medicine and healthcare". *Computers & Education* 40: 361-376.
4. Biggs J (1996). "Enhancing teaching through constructive alignment". *Higher Education* 32(3):347–64.
5. Botezatu M, Hult H, Tessma M, Fors U (2010). "As time goes by: stakeholder opinions on the implementation of a virtual patient simulation system". *Medical Teacher* 32(11): 509-516.
6. Botezatu M, Hult H, Tessma M, Fors U (2010). "Virtual Patient Simulation: knowledge gain or knowledge loss?". *Medical Teacher* 32(7): 562-8.
7. Botezatu M, Hult H, Tessma M, Fors U (2010). "Virtual Patient Simulation for learning and assessment: superior results in comparison with regular course exams". *Medical Teacher* 32(10): 845-850.
8. Bradley P (2006). "The history of simulation in medical education and possible future directions". *Medical Education* 40(3): 254-62.
9. Bryner BS, Saddawi-Konefka D, Gest TR (2008). "The impact of interactive, computerized educational modules on preclinical medical education". *Anatomical Sciences Education* 1(6): 247-51.
10. Colthart I, Bagnall G, Evans A, Allbutt H, Haig A, Illing J, McKinstry B (2008). "The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME Guide no. 10". *Medical Teacher* 30 (2):124-45.
11. Carroll JD, Messenger JC (2008). "Medical simulation: the new tool for training and skill assessment". *Perspectives in Biology and Medicine* 51(1): 47-60.
12. Cook DA (2005). "The research we still are not doing: an agenda for the study of computer - based learning". *Academic Medicine* 80(6): 541-8.
13. Cook DA (2006). "Where are we with Web-based learning in medical education?". *Medical Teacher* 28(7): 594-8.
14. Cook DA (2007). "Web-based learning: pros, cons and controversies". *Clinical Medicine* 7(1): 37-42.

15. Cook DA (2009). "The failure of e-learning research to inform educational practice, and what we can do about it". *Medical Teacher* 31(2):158-62.
16. Cook DA, Dupras DM (2004). "A practical guide to developing effective web-based learning". *Journal of General Internal Medicine* 19(6):698-707.
17. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM (2008). "Internet-based learning in the health professions: a meta-analysis". *JAMA* 300(10):1181-96.
18. Cook DA, McDonald FS (2008). "E-learning: is there anything special about the "E"?. *Perspectives in Biology and Medicine* 51(1):5-21.
19. Cook DA, Triola MM (2009). "Virtual patients: a critical literature review and proposed next steps". *Medical Education* 43(4):303-11.
20. Courteille O, Bergin R, Stockeld D, Ponzer S, Fors U. (2008). "The use of a virtual patient case in an OSCE-based exam - a pilot study". *Medical Teacher* 30(3): e66-76.
21. Creswell JW (2003). "Research design: Qualitative, quantitative, and mixed methods approaches". London, New Delhi: Sage.
22. Custers EJ (2010). "Long-term retention of basic science knowledge: a review study". *Advances in Health Science Education: Theory and Practice* 15(1):109-28.
23. Ellaway R, Masters K (2008). "AMEE Guide 32: e-Learning in medical education Part 1: Learning, teaching and assessment". *Medical Teacher* 30(5): 455-473.
24. Ellaway R, Poulton T, Fors U, McGee JB, Albright S (2008). "Building a virtual patient commons". *Medical Teacher* 30(2): 170-4.
25. Ellaway RH, Poulton T, Smothers V, Greene P (2009). "Virtual patients come of age". *Medical Teacher* 31(8):683-4.
26. Fall LH, Berman NB, Smith S, White CB, Woodhead JC, Olson AL (2005). "Multi-institutional development and utilisation of a computer-assisted learning programme for the paediatrics clerkship: the CLIPP project". *Academic Medicine* 80:847-55.
27. Feldman MJ, Barnett GO, Link DA, Coleman MA, Lowe JA, O'Rourke EJ (2006). "Evaluation of the Clinical Assessment project: a computer-based multimedia tool to assess problem-solving ability in medical students". *Pediatrics* 118(4): 1380-7.
28. Fors U, Muntean V, Botezatu M, Zary N (2009). "Cross-cultural use and development of Virtual Patients". *Medical Teacher* 31(8): 732-738.
29. Gesundheit N, Brutlag P, Youngblood P, Gunning WT, Zary N, Fors U (2009). "The use of virtual patients to assess the clinical skills and reasoning of medical students: initial insights on student acceptance". *Medical Teacher*, 31(8):739-42.
30. Haag M, Singer R, Bauch M, Heid J, Hess F, Leven FJ (2007). "Challenges and perspectives of computer-assisted instruction in medical education: lessons learned from 7 years of experience with the CAMPUS system". *Methods of Information in Medicine* 46:67-9.

31. Harden RM (2005). "A new vision for distance learning and continuing medical education." *The Journal of Continuing Education in the Health Professions* 25(1): 43-51.
32. Hawkins R, MacKrell Gaglione M, LaDuca T, Leung C, Sample L, Gliva-McConvey G, Liston W, De Champlain A, Ciccone A. (2004). "Assessment of patient management skills and clinical skills of practising doctors using computer-based case simulations and standardised patients." *Medical Education* 38(9): 958-68.
33. Huang G, Reynolds R, Chandler C (2007). "Virtual patient simulation at US and Canadian medical schools." *Academic Medicine* 82(5): 446-51.
34. Huwendiek S, Reichert F, Bosse HM, de Leng BA, van der Vleuten CP, Haag M, Hoffmann GF, Tönshoff B (2009). "Design principles for virtual patients: a focus group study among students". *Medical Education* 43(6):580-8.
35. Huwendiek S, de Leng BA (2010). "Virtual patient design and curricular integration evaluation kit". *Medical Teacher* 44 (5):519.
36. Issenberg SB (2006). "The scope of simulation-based healthcare education". *Simulation in Healthcare* 1:203–8.
37. Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ (2005). "Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review". *Medical Teacher* 27(1): 10-28.
38. Karpicke JD, Roediger HL (2008). "The critical importance of retrieval for learning". *Science* 319(5865): 966-8.
39. Kerfoot BP, Baker H, Jackson TL, Hulbert WC, Federman DD, Oates RD, DeWolf WC (2006). "A multi-institutional randomized controlled trial of adjuvant Web-based teaching to medical students". *Academic Medicine* 81 (3):224-30.
40. Kerfoot BP, Brotschi E (2009). "Online spaced education to teach urology to medical students: a multi-institutional randomized trial". *American Journal of Surgery* 197 (1):89-95.
41. Kneebone R, Bello F, Nestel D, Yadollahi F, Darzi A (2007). "Training and assessment of procedural skills in context using an Integrated Procedural Performance Instrument (IPPI)". *Studies in Health Technology and Informatics* 125: 229-31.
42. Kvale S (2007). "Doing interviews", 1st edition. Thousand Oaks, California: Sage Publications.
43. Kvale S (2009). "InterViews: learning the craft of qualitative research interviewing", 2nd edition. Los Angeles: Sage Publications.
44. Larsen DP, Butler AC, Roediger HL (2008). "Test-enhanced learning in medical education". *Medical Education* 42(10):959-66.
45. Ling Y, Swanson DB, Holtzman K, Bucak SD (2008). "Retention of basic science information by senior medical students." *Academic Medicine* 83 (10 Suppl):S82-5.

46. McLaughlin S, Fitch MT, Goyal DG, Hayden E, Kauh CY, Laack TA, Nowicki T, Okuda Y, Palm K, Pozner CN, Vozenilek J, Wang E, Gordon JA; SAEM Technology in Medical Education Committee and the Simulation Interest Group (2008). "Simulation in graduate medical education 2008: a review for emergency medicine". *Academic Emergency Medicine* 15(11):1117- 29.
47. Norman GR (2008). "The glass is a little full – of something: revisiting the issue of content specificity of problem-solving". *Medical Education* 42: 549-551.
48. Park J, MacRae H, Musselman LJ, Rossos P, Hamstra SJ, Wolman S, Reznick RK (2007). "Randomized controlled trial of virtual reality simulator training: transfer to live patients". *American Journal of Surgery* 194(2): 205-11.
49. Posel N, Fleiszer D, Shore BM (2009). "12 Tips: Guidelines for authoring virtual patient cases". *Medical Teacher* 31(8):701-8.
50. Roediger HL, Karpicke JD (2006). "Test-enhanced learning: taking memory tests improves long-term retention". *Psychological Science* 17 (3):249-55.
51. Rosen KR (2008). "The history of medical simulation". *Journal of Critical Care* 23(2): 157-66.
52. Round J, Conradi E, Poulton T. (2009). "Training staff to create simple interactive virtual patients: the impact on a medical and healthcare institution". *Medical Teacher* 8:764-9.
53. Schuwirth LW, van der Vleuten CP (2003). "The use of clinical simulations in assessment". *Medical Education* 37 Suppl 1: 65-71.
54. Sisti HM, Glass AL, Shors TJ (2007). "Neurogenesis and the spacing effect: learning over time enhances memory and the survival of new neurons". *Learning and Memory* 14 (5):368-75.
55. Steadman RH, Coates WC, Huang YM, Matevosian R, Larmon BR, McCullough L, Ariel D (2006). "Simulation-based training is superior to problem-based learning for the acquisition of critical assessment and management skills". *Critical Care Medicine* 34(1):151-7.
56. Steinert Y, Mann K, Centeno A, Dolmans D, Spencer J, Gelula M, Prideaux D (2006). "A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8". *Medical Teacher* 28(6):497-526.
57. Triola M, Feldman H, Kalet AL, Zabar S, Kachur EK, Gillespie C, Anderson M, Griesser C, Lipkin M (2006). "A randomized trial of teaching clinical skills using virtual and live standardized patients". *Journal of General Internal Medicine* 21(5): 424-9.
58. Turner MK, Simon SR, Facemyer KC, Newhall LM, Veach TL (2006). "Web-based learning versus standardized patients for teaching clinical diagnosis: a randomized, controlled, crossover trial". *Teaching and Learning in Medicine* 18(3): 208-14.

59. Tworek J, Coderre S, Wright B, McLaughlin K (2010). "Virtual Patients: ED-2 band-aid or valuable asset in the learning portfolio?" *Academic Medicine* 85(1):155-8.
60. Wahlgren CF, Edelbring S, Fors U, Hindbeck H, Ståhle M (2006). "Evaluation of an interactive case simulation system in dermatology and venereology for medical students". *BMC Medical Education* 6: 40.
61. Waldmann UM, Gulich MS, Zeitler HP (2008). "Virtual patients for assessing medical students - important aspects when considering the introduction of a new assessment format". *Medical Teacher* 30(1): 17-24.
62. Zary N, Johnson G, Boberg J, Fors U. (2006). "Development, implementation and pilot evaluation of a Web-based Virtual Patient Case Simulation environment - Web-SP". *BMC Medical Education* 6: 10.
63. Ziv A, Ben-David S, Ziv M (2005). "Simulation based medical education: an opportunity to learn from errors". *Medical Teacher* 27(3):193-9.