Author's response to reviews

Title: Virtual Patient Simulation: what do students make of it? A focus group study.

Authors:

Mihaela Botezatu (mihaela.botezatu@ki.se)
Håkan Hult (hakan.hult@ki.se)
Uno G Fors (uno.fors@ki.se)

Version: 2 Date: 6 October 2010

Author's response to reviews: see over
Reviewer 1

• Major Compulsory Revisions:

1 – 16 subjects are divided into two groups, but I can find no comparison of results between the groups. So, the Methods must be clarified.

No comparison was intended. We deliberately recruited more students than needed (according to literature, 9 is the maximum number of participants in a focus group), to allow students to decline participation or drop out at a later time. Equal numbers were recruited from the “study” and “control” arms of a previous study mentioned in the article* in order to avoid any biased opinions about the educational tool. When all students confirmed their participation, we had to arrange two focus groups, as 16 participants would have been too many to accommodate in one session. A clarification was added in the text.


2 – The Results seem to be authors’ a priori statements that have been 'fortified' by comments made by the participants. One would rather expect to codify the comments of participants and extract coherent and meaningful results from them.

The latter is precisely what we did (emerging coding). No a priori statements, judgements or pre-defined categories existed (i.e. the methodology was not a priori coding). This is now indicated in the text.

The value of reviewing videos for making the observations has not been established.

The videos were only reviewed in order to identify the voices on the tapes and assign a number to each participant, and this was indicated in the text.

Why no questionnaires?

The data offered by questionnaires are limited. We did use them in another study*; however, the aim of that study and its design were different, and therefore more suitable for questionnaires. We also wanted to allow for new issues (unknown to us, the researchers) to surface, and we think we were right to do that, judging by the results. We did mention our choice in discussion, although quite briefly; in the revised paper there is a more detailed consideration of method choice.


3 – The Themes and Categories of Fig.1 seem to represent concepts of the authors for which they use comments elucidated during focus group activities to validate those concepts.

No, they were derived from content analysis.

4 – This report would become effective if the five themes are identified as hypothetical topics about which focus group discussion was directed to get responses.

Even if this is a possibility, we did not intend to do so in our study.
A major emphasis on simulating real cases centers upon a photo of a patient that links to the scenario, and thus, direct transfer of virtual experience to medical practice. What if a different, a non-valid photo was substituted? Would the scenario no longer be relevant?

Indeed, that might be the case. Our experience from the classroom is that the students readily immerse in the proposed scenarios if they are reassured by proper media that the cases reflect their daily practice. Otherwise, they might dismiss the cases as non-relevant for their medical and socio-cultural context*. It may be argued that patient photos (all media related to a case for that matter) may be downloaded from repositories. This is true, but the university must first have access to such a repository and that is seldom the case; it is also a time-consuming job an overworked clinician is unwilling to undertake. We are not aware of any series of virtual patients built entirely on repositories; during their years of practice, clinicians collect media they later try to fit into a storyline. Regardless of origin of the media, the sex, race, age, location of signs in the physical exam and the location of lesions on radiographs, TAC, MRI etc must fit an otherwise credible scenario. Our choice was the collection of cases from the clinical environment in which the students learn and later practice as young professionals.


Reviewer 2

- **Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)**
  1. Consider a recent paper on authenticity in simulated patient work (Nestel & Kneebone, 2010, Academic Medicine). It is a different but relevant context. You might find it valuable.

Interesting context, dealing with the authenticity of standardized patient roles. We agree on the importance of acknowledging the existence of composite cases and on the need to investigate the educational benefits of more authentic vs. amalgamated cases. Our approach involved real clinical records and media, used to create credible storylines and feedback features for virtual cases.

- **Major Compulsory Revisions (which the author must respond to before a decision on publication can be reached)**
  1. I have mentioned copy editing above. As a start I would remove the terms booming, whopping, nowadays, febrile, and capitalize. I would ensure consistent and accurate use of the abbreviation VPS.

  Changed in the text.

  2. Focus groups are not ‘focus group interviews’. They are plain focus groups. This is an important distinction since group interviews with a focus are different to focus groups. See Morgan’s scholarly work on focus groups and group interviews.

  ‘Focus group interviews’ is a term that to our knowledge is used in a number of other publications and we would actually prefer it. As we see it, it is an accepted expression, in
Europe at least (where e.g. Steiner Kvale has contributed to the field with a number of works, such as ‘InterViews’ and ‘Doing Interviews’).

3. I’d like you to mention your finding in the discussion about communication and VPs. Although reported in results you do not mention this important point in the discussion.

 Changed in the text.

4. Although I recognise the difference between teaching and learning, the illustrations of the themes from the students’ perspectives was not clear to me. It seemed no different to their desired way to learn. Can you give some more thought to the need to separate these themes. Alternatively, present a different illustration.

 Learning and teaching, even if intimately intertwined, are essentially two different processes, with different stakeholders. In our analysis of the interviews we found opinions on both issues and subsequently reported them as separate themes.

5. In the abstract it would be helpful to make clear in the results that you identified 5 themes rather than read to the conclusion. Provide the reader with a framework first and then fill in the gaps.

 Changed in the text.

6. I found the reference to who did what (that is, the research team members) in the text confusing. Please use the same system for referring to authors either by initials or some other consistent means (in sections 2, 3 & 4).

 Changed in the text.

7. Even though the ‘trusting’ relationship was established with one of the researchers and students, s/he was still in a relative position of power. I don’t think this can be dismissed, especially if the students knew that this person was ‘committed’ to this project. It just needs to be acknowledged. Because no field notes were taken it is hard to report whether there was a perceived influence or not.

 Changed (acknowledged) in the text.

8. I also found the layout of the verbatim statements very hard to read within the mixed text of italics and non-italicised text. It is a small but important point to separate the italics by a line. If you are just presenting a few words or the odd sentence, then this current approach would be fine but several paragraphs were difficult to read. It is probably best to use conventions in presenting data.

 Actually, we even considered the possibility of providing a separate table with all the quotes and one of the early versions did include a table. We found that to be more confusing than the present layout, as the reader was forced to turn the page and look for the relation between the quote and the text. However we modified the layout and we hope it is now friendlier to the reader.

 Changed in the text.

9. I don’t think you should use the patient’s name, even with permission. It is unnecessary. (Perhaps it was a pseudonym but it adds nothing).
We agree, deleted.

10. Start the discussion with the five themes rather than a complicated weaving of themes.

We think that the initial paragraph, which was rewritten, motivates our choice of focus groups. The rest of the Discussion has the same stepwise layout as the Results.

11. North American ED2 is jargon. It did not mean anything to me. HEI’s also needs spelling out. Perhaps I missed it.

Replaced / spelled out.

12. In the discussion section on learning, you presented a series of sentences not in paragraph form so this was confusing too.

It is a discussion of the categories, which are italicized in order to be more easily discerned from the rest of the text. We are not sure we completely understand this question / suggestion.

13. Review your language on assessment and examinations and consider the use of formative and summative assessment. I was not always clear what you (or the students) were referring to.

We have tried to make this clearer in the text, where appropriate.

14. Ethics

A sentence was added in Method.

Reviewer 3

Research questions

It would add to the value of the article if the specific characteristics of this VSP were related to other relevant research. The assumptions of the relation between user expectations and success/failure of the intervention seem relevant but need to be related to the research field at large.

A search of major databases using the criteria virtual patient was commissioned to the Library of Karolinska Institutet in May 2010 for the benefit of my doctoral thesis (the ‘kappa’). 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 truly be about VPS; the rest dealt with virtual worlds, virtual reality, computational modeling 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 e.g. assessment, out of a pool 232 articles published in 2005.
It is therefore difficult to establish a relationship with the VPS research field (and even with the more generous field of simulation), as relevant articles are missing from the literature. On the other side, the field of medical education would be too broad, hence out of the scope of the present research.

Methods

What problem/question (if any) is introduced?

Several problems were introduced, as the students had used the system for learning and / or for assessment. They had also collected clinical cases in the hospitals. So the discussion began by recapitulating their experience with the system and by stating the scope of the interviews, i.e. the exploration of their perception of the system. Examples of questions are: Is the system good for learning? Why? In what way? What do you think about the features of the system – all those that are needed? Missing? Supporting learning? If yes, how? What do you think about the quality of the feedback? etc. At times the students themselves raised issues we did not plan to bring up, such as communication skills development or their motivation in using the system.

Is the researcher active?

Yes. In Method we indicated that MB led the interviews.

Given 8 students per group and a discussion time of about an hour, each student had in average about 7 minutes to express her/his opinion and give “a richer and more detailed” contribution. Who talks and who is quiet? Somewhat longer transcripts from the discussion in order to answer such questions should be included.

The way we presented the results is probably misleading with regards to the broad student participation during the interviews. The ‘best’ quote was chosen (the most articulate); this does not mean that other students did not express the same idea in words (and not merely agree). We agree that a ‘quantitative’ presentation of results is probably not the best way to go about; we did consider enumerating all the identification numbers for all the students who actually said very similar things, and they turned up to be quite many. An early version had all the ID numbers and at first glance it was quite difficult to discern whether the numbers were students or references. So for the sake of clarity we opted for expressions such as ‘some’, ‘many’, ‘the majority’ etc., as it was the only way to convey the meaning without boring the reader. Longer transcripts may not serve the cause either, because the information becomes redundant and the reader may not find any added value in reading them.

The paper must, further, include a discussion of the relevance of the method chosen in relation to the aim and research question. Are 16 of 216 students enough to satisfy the interest to gather a wide range of opinions? The article ought to discuss the question.

We intended to collect as much data as possible while still keeping the project feasible. After discussing the possibility of using in-depth interviews and questionnaires as opposed to focus groups, we chose the latter. We felt that while still allowing the collection of sufficient, good quality data, it was also a practical way (financially and time-wise) to investigate the participants’ opinions. Students have actually come up with descriptions, explorations and ideas that might have not surfaced in individual interview. Obviously, we do not pretend to
have captured all the opinions that may exist in the larger population (and of course it is not necessary to interview a whole population in order to get most of the data). Simple randomization was also means of ensuring that the results were not biased.

The merits of focus group interviews compared to individual interviews should thus be discussed, as well as the value that the focus group interview can add to questionnaires.

A short clarification was made in the beginning of Discussion.

Results
The results are presented as five themes with altogether 18 categories. Nothing is said of how the themes and categories are constructed.

We agree and we added a section in Method (dealing with emerging coding, among others).

The 7 minutes that each student on the average had to her/his disposal to discuss, which led to the forming of 18 categories, probably meant that many of the students only had a chance to talk on a few of the topics. As more than half of the quotes (18 of 34) are quoted from 4 students, the question is if the aim of capitalizing on the variation is taken care of. The number of quotes in relation to the number of students actualizes the question of choosing focus group interview as a research method. This ought to be discussed.

As previously discussed, we acknowledge the issue. However, we would like to stress that even if it may seem that half of the results rely on four students, this is not what happened. We simply chose not to list all the ID numbers and the redundant quotes, after a thorough discussion of the first drafts of the paper.

Coming back to the reason for adopting a qualitative method - to focus on the variation of students’ opinions and experiences, the way the result is presented is confusing. The results are frequently presented not with a focus of the variation of the opinions expressed on a specific issue, but as the opinion of the majority of the students: “the participants greatly appreciate”, “most students consider”, “students think”. Such comments are not in line with the character and goal of qualitative analysis but lends from quantitative analysis.

Please see above.

The discussion of the results actualizes a number of the problems pointed to above. The fact that the qualitative approach, according to the authors, resulted in results similar to the quantitative research presented earlier (p. 17) needs a comment. In what ways did the work at hand add to the results of the larger, quantitative study? Did the focus interview clarify or deepen the understanding that was the result of the larger study? The article mentions some results that add to what was known before, but so briefly that it is difficult to understand (eg p17 Learning). For example, what is added by the comment “Medical students feel they remember more with VPS” (p 17) to the retention enhancement demonstrated by the larger study? As the data is treated as quantitative rather than as qualitative, the article largely fails to demonstrate the merits of applying a qualitative method. It is difficult to see “the wider range of opinions” or the “richer and more detailed end results” that are promised.

We rewrote the first paragraphs in discussion. The sentence in question was eliminated, as well as the reference to quantitative research.
Instead, generalizations, to which the data does not lend itself, are frequent. An example is the conclusion “We believe our finding could be generalized to VPS of similar design” (page 20), a conclusion that would have great importance, but which can hardly be drawn based on what is presented in the article.

We do have a very good understanding and own experience with similar systems with a linear design. The most recent example is the eVIP project of the EU, in which Karolinska was a member. Our study did not aim to compare systems with linear design. Moreover, that would be of no value to the VPS field: they do work similarly, the difference being that some provide feedback (such as Web-SP) and others do not. Based on our in-depth knowledge of other systems we do believe that these results can be generalized to all applications of similar design. The generalization in question was located in the section on ‘limitations’ and was flagged up by the verb ‘believe’. However, further clarifications were added in the text.

The discussion does refer to earlier research of VSP used in medical education, but it lacks references to the well established and relevant research field of teaching, learning and assessment. Researchers have since long studied the merits of, for example, authenticity as a tool for learning and assessment, and have also qualified the context of authenticity in line with the comments made by some of the students who are quoted. The same is true for the transparency of examination goals and for feedback.

The article is circumscribed to medical education, simulation and more specifically to VPS simulation. Our target readers are doctors working with similar systems. We would therefore like to keep our discussion contained within the field the readers are comfortable in.

The reviewed work is part of a larger study “on virtual patient assessment results” (p 6). It is difficult to see how the actual work is related to the larger study as this study is only mentioned but not described. The lack of a clear research question adds to the difficulty. I have, however, understood that both studies focus on performance assessment with inbuilt feedback. As this type of assessment has attracted much research attention also outside medical research, it would be of value to anchor the work in such research. Researchers have, for example demonstrated a strong link between feedback and self-assessment for dental students in an interactive examination using virtual patient simulations. The problem of assessing complex performance in a credible way is another topic of relevance which could have been referred to, especially so, as students express positive feelings about the authenticity of the examination. Reviews have pointed to the importance of “transparency” in strengthening the learning effects of performance assessment. My conclusion is, thus, that the article would benefit from being located in a relevant research context.

The population is part of a larger cohort. The current study is not a part of a larger study. The issue of the relevant research context was discussed earlier.