

Does High Fidelity Make the Music Sound Better? Relative Effectiveness of High vs. Low Fidelity Simulation in Learning Heart Sounds

Daniela de Giovanni¹, Trudy Roberts¹, and Geoff Norman²
University of Leeds (1) and McMaster University (2)

A number of recent review articles have extolled the virtues of high fidelity patient simulation for learning many skills related to clinical medicine, from relatively simple skills such as recognition of heart sounds to complex skills of crisis management^{1 2 3}. Some enthusiasm appears justified; high fidelity simulations have a number of potential advantages over actual patient experiences – the potential to provide standardized and graded experiences, reduction in use of faculty instructor time, opportunity for contact with rare or life-threatening situations in a low-risk environment. Further, it would seem that the advent of high fidelity simulations that include many aspects of a real patient represent a genuine advantage over less realistic simulations. Schuwirth and van der Vleuten⁴ begin with the premise that “authenticity should have high priority when programs for the assessment of medical competence are being designed... the situations in which a candidate’s competence is assessed should resemble the situation in which the competence will actually have to be used”. Such assumptions are consistent with some models of learning such as Miller’s pyramid⁵, in which the apex of the pyramid is “Does” and the base is “Knows”, implying again that the closer we can get to actual performance, the better.

Some evidence suggests that this may be true. One widely cited study⁶ showed that students who learned to recognize heart sounds on a high fidelity simulator showed a 33% gain in diagnostic accuracy on recognizing simulated heart sounds after a 2 week elective, compared to a 5% gain in a control group who had usual bedside teaching. An older study⁷ showed less impressive, but still significant gains of 19% in a simulator trained group vs 6% for a control group who had usual teaching.

However, these gains may be somewhat optimistic for two reasons. First, in both studies, the comparison was between the controlled instruction on the simulator and “bedside” teaching; no attempt was made to determine how much teaching actually occurred at the bedside. Indeed, given the rarity of abnormal heart sounds, there is a possibility that the control students actually had no exposure to any abnormalities. Second, in both studies, the outcome was assessed on the same simulator that the students in the experimental group learned on. Thus, they need not demonstrate the ability to recognize mitral stenosis or aortic regurgitation; they need only recognize the particular example of mitral stenosis that was encountered during instruction. Hence, although one can argue that high fidelity simulation, because it is so realistic, should lead to good transfer to the real world, these findings do not demonstrate transfer. However, the Ewy study examined performance with real patients and found only a 2.8% difference between the two groups.

There is a third issue that is not addressed in the studies to date. Although it may be the case that high-fidelity simulation has some advantages over alternatives, this is achieved at some cost. “Harvey”, the heart sound simulator, costs about \$75,000, effectively ensuring that no school can afford more than one. If it can be shown that a lower fidelity simulator (such as a CD) can achieve similar learning gains at substantially lower cost, then potentially each student could have considerably more time to learn the materials. We are not the first to raise the issue of the relation, or lack of, between fidelity and learning. Kneebone writes, “All too often it is the surface realism of the simulation that occupies the ingenuity of those who develop it, eclipsing key issues of teaching and learning... lower levels of fidelity may reduce technological limitations and cost without compromising outcomes”⁸.

A few studies have experimentally compared low and high fidelity instruction, and, perhaps surprisingly, Kneebone’s cautionary note appears well grounded. Two studies have examined low vs. high fidelity simulation in learning surgical skills^{9 10}, and both found that the low fidelity simulator had slightly, but not significantly, smaller gains than the high fidelity group. Gilbert¹¹ compared a seminar in trauma management to a computer-based simulator, and to two control groups. Both instructional groups showed large gains compared to control, but there was no significant difference between seminar and simulation. Finally, Wenk¹² showed that a problem-based discussion was as effective as a simulation in learning anaesthesia skills.

There is one final consideration relating to an earlier point. Any physical sign can be manifest in myriad different forms. Part of learning data interpretation is recognizing the many ways in which a sign may present. Testing on the same example as instruction may overestimate the effect of learning, as shown in the Ewy study. Conversely, a high fidelity simulation such as Harvey, with only one example of each condition exemplifies Kneebone’s concern that physical accuracy puts learning issues in the back seat. It may well be that a far lower fidelity simulation (i.e. a CD) with multiple examples of each condition, may actually show superior performance to the high fidelity – one example simulator, when tested on novel examples.

The present study was designed to address these issues. We conducted an experiment wherein, after some basic instruction in auscultation, students learned some abnormal heart sounds using either a) Harvey or b) a CD. Six weeks later, they were then tested in an OSCE format with a) Harvey sounds, b) CD sounds and c) Real patients with stable heart abnormalities. This study was preceded by a small pilot study to validate the instruction and the testing materials, where the outcome was interpretation of old and new heart sounds on Harvey and CD immediately after instruction.

Research Hypotheses:

a) Students who learn from Harvey will show better learning than students who learn from a CD in a) clinical process, b) recognition of specific heart sounds, c) auscultatory diagnosis, when assessed with real patients.

b) Students who learn heart sounds from one modality (Harvey or CD) will show higher recognition for heart sounds learned in training on that modality

Pilot Study – Immediate Ability and Near Transfer

Methodology

Sample

10 second year medical students at the University of Leeds volunteered to take part in the pilot study. Students were randomly allocated to CD or Harvey group by picking colored sweets which were then assigned to each group.

Instructional Procedures

The teaching and testing occurred on the same day. Students received an introductory lecture outlining the structure of the trial and refreshing basic cardiovascular examination technique. The groups were then divided and taken through the teaching material. The content of the tutorials was similar to the main study however ventricular septal defect and mechanical heart valves were omitted so as not to overload the students with information. Students received 3 one hour tutorials with a 5 minute interval between sessions 1 and 2 and a 30 minute lunch break between sessions 2 and 3. Session 1 covered normal heart sounds, session 2 systolic murmurs (aortic stenosis and mitral regurgitation) and session 3 diastolic murmurs (aortic regurgitation and mitral stenosis).

The groups were taught by the same instructor for each of the 3 tutorials. The instructors were clinical educators with experience in teaching cardiovascular examination.

At the end of the final session 3 practice examples were given on the allocated teaching tool as a mock test. Students familiarized themselves with the mark sheets and received feedback if incorrect answers were given.

CD

The 5 students in this group were taught together. The lesson covered the basic principles behind each condition, the expected auscultatory findings and associated clinical signs. Examples of the relevant condition were then broadcast using a laptop computer. There were 21 examples in total: 4 normal heart sounds; 4 aortic stenosis; 6 mitral regurgitation; 3 aortic regurgitation and 4 mitral stenosis. Each example was played 3 times to give students the opportunity to appreciate the individual features of each example. It was also felt that this was reflective of the time Harvey students had to listen to the different conditions.

At the end of the final session 3 examples were broadcast in a mock test. The students were given feedback on their answers with explanation where possible if murmurs were confused.

Harvey

The 5 students in this group were taught together. The same lesson plans were used as the CD group and after each condition was taught, students would listen to just 1 setting on Harvey. One student would examine Harvey with the stethoscope whilst the remaining students listened with the headphones. The students then rotated round so that each student examined Harvey for each condition.

At the end of the final session the same 3 conditions, but with Harvey's example, were used in a mock test. One student examined Harvey and the remainder listened with the headphones. The answers were discussed and feedback given.

Measurement

All students were tested together at the end of the final session. 8 examples were broadcast from the laptop (5 examples had previously been played to the CD group; 3 examples were new to all students) and 7 examples were broadcast using speakers attached to Harvey (again 5 examples had previously been used with the Harvey group and 2 examples were new). Each example was played twice and lasted approximately 30 seconds. From the CD examples 1 old example of each of the 5 taught conditions was broadcast, followed by a new example of aortic stenosis, mitral stenosis and normal heart sounds. The 5 settings used in the teaching sessions on Harvey were broadcast followed by new examples of mitral regurgitation and aortic regurgitation.

After listening to each simulated heart sound, the student recorded their impression by indicating on a checklist, a) what heart sound(s) they heard, and b) the diagnosis. These were subsequently scored as correct or incorrect.

Analysis

Overall correct scores were created for each student for 3 conditions – Total score on Old Harvey sounds, Total score on Old CD sounds, and Total score on New sounds (Harvey and CD). The expectation was that students who had learned on the CD would do better on the CD and vice versa, as an indication of transfer specificity. The comparison of Harvey and CD old sounds was conducted using a repeated measures ANOVA, with the critical test being a significant interaction between Harvey / CD training and Harvey /CD sounds. Transfer to new items was tested using a t test on the means of the 2 groups

Results:

The results are shown in Table 1:

Table 1 about here

Considering first the Old sounds (which were, of course, only Old for one or the other group), the Harvey group performed as expected, obtaining an average score of 72% when tested with sounds they had heard during training, and dropping to 36% with CD sounds. The CD group, however, showed no drop in performance when encountering Harvey sounds (60% vs. 60%). Examination of transfer to new examples indicated no difference between the groups (CD = 48%, Harvey = 44%, $F = .05$, $p = .82$).

Discussion

Although the results were not statistically significant, the trends were in the anticipated direction. In particular, it appeared there was modality-specific performance, with Harvey-trained students doing best on Harvey and vice versa. However, one surprise is that the CD group showed no decrement in performance when examined with the Harvey sounds. The fact that the two means are identical must be viewed as coincidental, since the means are actually based on different stimuli, and it may well be that the Harvey sounds, were on average somewhat easier. Nevertheless, there is some basis for expecting that the CD trained group may well do worse on Old examples, but generalized better. Since they had seen multiple examples of each condition, it is not surprising that, on the one hand, individual learning examples are not as memorable, and on the other, that they are able to generalize to new examples, so cannot be viewed as definitive. ??better. Nevertheless, this advantage did not appear to hold with the New examples the pilot study did serve to validate the materials and procedures, and so we then proceeded on to the main study.

Main Study – Transfer to Real Patients after Delay

This study was intended to address the critical question of the extent to which training on a high or low fidelity simulator results in acquisition and retention of auscultation skills that can be effectively applied to real patients.

Methodology

Sample

The sample was 37 medical students from the University of Leeds in their 3rd year. The experiment took part at the beginning of the academic year and participating students were considered equally naïve in cardiovascular auscultation. All students were volunteers. Students signed up to Group A or B and were then group randomized to Harvey or CD tuition.

Instructional Procedures

All students received an hour long introductory lecture covering the same aspects as in the pilot study, this again lasted 1 hour. All the students were then introduced to Harvey to avoid bias in the final assessment. Students attended 3 further 1 hour tutorials with standardized lesson plans in groups of 6. In total students received 4 hours of tuition,

which is in keeping with the usual practice of the department when teaching cardiovascular examination. The first session covered normal heart sounds, including splitting of the second heart sound, and aortic stenosis. The second session included mitral regurgitation, ventricular septal defects and aortic regurgitation. The final session covered mitral stenosis and mechanical heart sounds. Neither group heard examples of the mechanical heart sounds, but the principles were taught to both groups.

The teaching material for both groups was identical. Each condition was approached by discussion of the auscultation findings focusing on features distinguishing between similar murmurs. Students were then taken through the other clinical signs that might be encountered before listening to their designated examples.

The instructors varied between sessions and taught to both groups to eliminate bias. The instructors were clinical educators with experience of teaching cardiovascular examination.

At the end of the final session 5 practice examples were given on the allocated teaching tool as a mock test. Students familiarized themselves with the mark sheets and received feedback if incorrect answers were given.

CD

The CD group were taught in groups of no more than 6 students. They received 3 tutorials each lasting 1 hour, roughly 1 week apart. For each condition students were taught the basic principles behind the auscultatory findings, what they should be hearing from the examples and the associated signs of that particular condition. The relevant examples were then broadcast highlighting the individual qualities of each. Each example was played once and lasted on average 30 seconds. In total 33 examples were broadcast – 7 normal heart sounds; 6 Aortic Stenosis; 6 Mitral regurgitation; 6 Mitral stenosis; 4 aortic regurgitation and 2 ventricular septal defect. Students were encouraged to ask questions which were appropriately answered at any point in each session.

At the end of the final session 5 examples were broadcast in a mock test. The students were given feedback on their answers with explanation where possible if murmurs were confused.

Harvey

Students were also taught in groups of no more than 6 students and received 3 hour long tutorials each separated by 1 week. The same lesson plans were used. However after each condition and associated signs were described, students used Harvey to gain experience of the heart sounds. Each student examined Harvey using the stethoscope and the remaining students listened with headphones. Students were allowed to palpate and time the heart sounds if they wished to. The students then rotated round so that each student had the opportunity to examine Harvey with each condition. Questions were again encouraged at any point in the tutorials.

At the end of the final session the same 5 conditions were used in a mock test using Harvey's examples. One student examined Harvey and the remainder listened with the headphones. The answers were discussed and feedback given.

Measurement

All students in both groups underwent a 6 station OSCE approximately 6 weeks after the initial instructional session. The OSCE contained stations with real patients, and one station with simulated sounds.

Patients

A total of 8 patients were recruited. One had a mechanical heart valve, 3 had aortic stenosis, 1 with aortic and mitral regurgitation, 1 with mitral regurgitation, and 1 with normal heart sounds. While the sample of conditions displayed by real patients is restricted to conditions which can be left untreated, it is notable that all the conditions except the mechanical heart valve are ones in which the Harvey group, which had learned to synchronize heart sounds with the carotid pulse, should have an advantage.

Each student saw 5 patients; to reduce fatigue on patients, individual patients saw from 10 to 30 students. Each patient encounter lasted 10 minutes. Each patient encounter was observed by a clinician, who provided feedback on performance using 5 point scales rating a) Communication skills and b) Examination skills from 1 = Clear Fail to 5 = Excellent Pass. In addition, at a last station, all students listened to 3 heart sounds on a digital recorder, chosen from the CD examples and examined Harvey with 3 different example settings. The 6 simulated heart sounds were, from the CD, normal heart sounds, aortic stenosis and aortic regurgitation, and from Harvey, mitral stenosis, ventricular septal defect, and aortic stenosis. All examples had been previously heard by the instructional group using that medium, as part of their instruction.

After each patient and after listening to each simulated heart sound, the student recorded their impression by indicating on a checklist, a) what heart sound(s) they heard, and b) what was their diagnosis. These were subsequently scored as correct or incorrect.

Analysis

An overall analysis was conducted by creating a total score for each student on: CD diagnosis correct /3, Harvey diagnosis correct / 3, Correct murmurs from CD and Harvey / 6, Correct murmur from the patient, Correct diagnosis from the patient, and examiner feedback scores for communication and examination, averaged across the 5 patients. Overall means for the two groups were compared with a t test.

Results

The results are shown in Table 2.

Table 2 about here

Apparently, training with Harvey resulted in a small, but not quite significant, advantage in interpreting clinical signs of real patients on auscultation, but the diagnostic accuracy for the two groups was very close. Moreover, there was no evidence that direct “hands-on” training with Harvey resulted in improved clinical skills as assessed by an observer, either in communication skills or, more critically, in examination skills.

One possible explanation for the non-differences may be confounding. Because each student saw only 5 of the 8 patients, it is possible that patients were distributed across the two learning groups in an unbalanced fashion. To examine this more closely, we looked at differences between the groups patient-by-patient. For each patient, we computed the mean score for accuracy of murmurs and accuracy of diagnosis for all the students who examined the patient. We then calculated a t test on the differences for the 8 cases.

The results are shown in table 3 below:

Table 3 about here

The results are substantially the same as the first analysis. The Harvey group appeared somewhat, but not significantly, better in interpretation of murmurs, but there was no difference in diagnostic accuracy.

Discussion

The results of the main study are reasonably unequivocal. The two groups who had equivalent time of instruction using actual contact with the Harvey mannekin and listening to a CD were virtually equivalent in performance on the real patients. There was some suggestion that the Harvey-trained group were better at interpretation of heart sounds, but the result was not significant. Conversely there was no evidence that the Harvey trained group were better at diagnosis, nor was there any evidence that their observed clinical and communication skills were any better.

One possibility is that in fact, the six week delay resulted in such a loss in learning that the two groups were back to baseline. There are two arguments against this: first, the small sample of students in the pilot study was at about the same performance level, 50%, immediately after instruction. Second, it would be somewhat unfortunate if instruction had a half life of only six weeks.

There may not have been sufficient power to detect a difference. The sample sizes are small, however we have calculated that the study would be able to detect a 20% difference in diagnostic accuracy (2.84 → 3.40) with power of about .65. Absolute gains

of 33% in the Harvey group, as observed by Issenberg et al (1999) would be detected with 98% power.

The findings are also interesting in that there was no evident difference in the observed process of clinical skill, despite the opportunity for hands-on examination for students in the Harvey group. It may be that relatively little instruction is necessary to achieve a satisfactory observed process. On the other hand, we did not have a chance to subject the form to reliability and validity assessment; it may be that it is not sufficiently reliable. A second surprise that the stable conditions used in the OSCE, mitral regurgitation, mitral stenosis, aortic regurgitation, aortic stenosis are precisely those where Harvey, with the ability to coordinate carotid pulse with auscultation sounds, should have a clear advantage. It does seem that it has a small advantage in labeling abnormalities, but this does not translate into improved diagnostic accuracy.

General Discussion

The present studies have overcome two methodological problems inherent in most studies to date of high fidelity simulation for auscultation training. First, we have experimentally contrasted high fidelity training with a more economical low fidelity condition, to determine what advantage is derived from the presence of the high fidelity simulator. Second, we have explicitly examined transfer, both to other heart sounds portrayed on the simulator and to real patients. As it turns out, there is an apparent advantage of simulator training, when it comes to recognizing heart sounds *on the simulator*. The small pilot study showed that students trained on Harvey showed an improved accuracy of 36% when tested with the same heart sounds, very similar to the 33% improvement documented by Issenberg¹². But we have shown that this advantage is not sustained when tested by new heart sounds, either on a simulator or on real patients. While Harvey trained students showed overall accuracy with real patients of about 50%, equivalent performance was achieved by the control group using a CD. Further, there was no measurable advantage in clinical skills from Harvey-based training.

As we described earlier, the findings are consistent with a growing literature in the simulation field showing that, while high fidelity simulators may be a “gold standard” in the absence of real patients, nevertheless, nearly as much may be gained with “low-fidelity” simulators at greatly reduced cost. Since an educational truism is that amount learned is directly related to time spent learning, there may well be an explicit tradeoff between the fidelity (and cost) of the simulator and the amount of time any student can access it.

The study also underlines the point raised by Kneebone and others (ref) that high fidelity is not necessarily better than low fidelity. Instead, effective use of simulation demands a careful analysis to determine the critical task demands, and identify which elements of the simulation must be present to ensure effective transfer to real patients.

Table 1
Pilot Study. Performance of Harvey and CD trained groups on Old Harvey sounds, Old
CD sounds, and New sounds (Percent correct)

GROUP	Harvey (k=7)	CD (k=8)	New (k=5)
Harvey	.72 (.27)	.36 (.12)	
CD	.60 (.27)	.60 (.12)	

Table 2
Outcomes on Harvey, CD and Patients

	Harvey Group	CD Group	
	Mean (SD)	Mean (SD)	t (p)
Harvey / CD			
Correct murmur	2.56 (.92)	2.53 (1.26)	.08 (.94)
CD Diagnosis	1.00 (.84)	0.79 (.71)	0.82 (.41)
Harvey diagnosis	1.33 (0.59)	1.16 (0.90)	.70 (.49)
PATIENT			
Correct murmur	3.11 (.90)	2.47 (1.12)	1.89 (.06)
Correct Diagnosis (2.94 (1.0)	2.84 (1.06)	.30 (.76)
Obs Feedback Comm Skills	18.89 (2.72)	19.68 (2.67)	-.89 (.37)
Obs Feedback Examination Skills	17.44 (2.61)	17.53 (2.63)	-.09 (.92)

Table 3
Results grouped by Patient

	Mean (SD) Harvey	Mean (SD) CD	t (p)
Murmur	0.56 (0.33)	0.49 (0.15)	0.57 (0.58)
Diagnosis	0.53 (0.29)	0.52 (0.16)	0.07 (0.94)

¹ Issenberg SB, McGaghie WC, Petrus ER, Gordon DL, Scalese RJ. Features and use of high fidelity medical simulations that lead to effective learning: A BEME systematic review. *Med Teach* 2005; 27: 10-28

² Maran NJ, Glavin RJ. Low to high fidelity simulation – a continuum of medical education? *Med Educ* 2003; 37(Supp): 22-28.

³ Bradley P. The history of simulation in medical education and possible future directions. *Med Educ* 2006; 40: 254-262

⁴ Schuwirth L, van der Vleuten CPM. The use of clinical simulation in assessment. *Med Educ* 2003; 37(Supp): 65-71.

⁵ Miller GE. The assessment of clinical skills/competence/performance *Acad Med* 1990; 65: S63-7.

⁶ Issenberg SB, Petrus ER, McGaghie WC, Felner JM, Waugh RA, Nash IS, Hart IR. Effectiveness of a computer-based system to teach bedside cardiology. *Acad Med* 1999; 10(supp): S93-95.

⁷ Ewy GA, Felner JM, Juul D, Mayer JW, Sajkid AW, Waugh RA. Test of a cardiology patient simulator with students in fourth year electives. *J Med Educ* 1987; 62:738-43.

⁸ Kneebone R. Evaluating clinical simulations for learning procedural skills: a theory-based approach. *Acad Med* 2005; 80: 549-553.

⁹ Anastakis DJ, Regehr G, Reznick RK, Cusimano M, Murnaghan J, Brown M et al. Assessment of technical skills transfer from the bench training model to the human model. *Am J Surg* 1999; 177:167-170.

¹⁰ Matsumoto ED, Hamstra SJ, Radomski SB, Cusimano MD. The effect of bench model fidelity on endourological skills: a randomized controlled study. *J Urol* 2002; 167:1243-1247.

¹¹ Gilbert MK, Hitchison CR, Cusimano MD, Regehr G. A computer-based trauma simulator for teaching trauma management skills. *Am J Surg* 2000; 179: 223-8.

¹² Wenk M, Waurick R, CSchots D, Wenk M, Gerdes C, Van Aken HK, Popping DM. Simulation-based medical education is no better than problem-based discussion and induces misjudgment in self-assessment. *Adv Health Sci Educ*