Long-term Retention of a 3-Dimensional Educational Computer Model of the Larynx

A Follow-up Study

Dieter Fritz, MD; Amanda Hu, MD; Tim Wilson, PhD; Hanif Ladak, PhD; Peter Haase, PhD; Kevin Fung, MD, FRCSC

Objectives: To determine the long-term retention of a 3-dimentional (3-D) educational computer model of the larynx to teach laryngeal anatomy and to compare it with standard written instruction (SWI).

Design: Prospective randomized controlled trial.

Setting: University education program.

Participants: One hundred health care students.

Interventions: For short-term assessment, 50 students were randomized to the 3-D model and 50 to SWI and were tested using a 20-question laryngeal test. Six months later, the same students were invited to retake the laryngeal anatomy test to examine long-term retention.

Main Outcome Measure: The score on a 20-item Web-based test that assessed the students’ level of knowledge of laryngeal anatomy approximately 6 months after their initial exposure to the laryngeal anatomy teaching intervention.

Results: Sixty-two students retook the test: 3-D (n=30) and SWI (n=32). No significant difference was noted in mean scores (P=.54) and change in scores (P=.59) between short- and long-term retention on the laryngeal anatomy test. There was a trend toward an increase in 3-D scores in both groups (P=.07) and a significant increase in 3-D scores in the 3-D group only (P=.049).

Conclusions: A low-fidelity model (SWI) is just as effective as a high-fidelity model (3-D) in teaching laryngeal anatomy. The acquired knowledge from either educational intervention may last up to 6 months for long-term retention. This study is one of the few in medical education to examine long-term retention.


Undergraduate medical education continues to evolve as the challenge of optimizing its content and its delivery is ever increasing. The teaching of human anatomy has been a pillar of medical education since its inception; however, in recent years, the emphasis on teaching gross anatomy and the allotted hours in the curriculum have been steadily declining.1-3 Reasons for this shift in curriculum are multifactorial and center on the issues of costs associated with the maintenance of cadaver dissection laboratories, shortages of qualified instructors to teach human anatomy, ethical considerations regarding cadavers, increased medical school enrollment, an increase in the amount of curriculum content, and the trend toward increased use of distributed medical education.4

As a result of these limitations, there is a need for alternative teaching modalities for human anatomy. Computer-assisted instruction (CAI) models have shown promise in the circumvention of these limitations. Studies5-7 have shown that when learning difficult concepts, CAI models are more efficient, are less stressful, and have equal or better effectiveness compared with traditional didactic lecture instruction. For example, it is thought that one of the main strengths of traditional cadaver dissection is the facilitation of learning spatial relationships among anatomical structures by providing 3-dimensional (3-D) views that cannot otherwise be represented in 2-dimensional (2-D) pictures. With the development of new 3-D CAI models, this weakness in past CAI models may no longer be a factor. Another strength of CAI models is student preference; a previous study8 noted that students found the 3-D CAI model to be more enjoyable than standard written instruction (SWI). Student preference is an important factor because it aids in the increased use of teaching modalities.9

Author Affiliations:
Departments of Otolaryngology–Head and Neck Surgery (Drs Fritz, Hu, and Fung), Anatomy and Cell Biology (Drs Wilson and Haase), and Medical Biophysics and Electrical and Computer Engineering (Dr Ladak), University of Western Ontario, London, Ontario, Canada.

©2011 American Medical Association. All rights reserved.
We previously created an interactive Web-based 3-D computer-learning model of the larynx1 and tested its efficacy in teaching anatomy.8 Students believed that this model was a preferred and valuable supplement to traditional teaching methods.8 However, long-term retention of laryngeal anatomy from this learning model is unknown. The present study is a 6-month follow-up to the previous prospective randomized controlled trial. The study objectives were to determine the long-term retention of the 3-D computer model to teach laryngeal anatomy and to compare the long-term retention of a 3-D educational computer model with SWI.

METHODS

This study was a prospective randomized controlled trial using a 3-D educational model of the larynx. A detailed description of the development of the model can be found in a previous study.1 We used each student’s score on a 20-item Web-based test to assess the student’s level of knowledge of laryngeal anatomy approximately 6 months after his or her initial exposure to the laryngeal anatomy teaching intervention. The test consisted of 13 factual questions and seven 3-D questions that evaluated each participant’s understanding of 3-D spatial relationships or his or her ability to identify a structure from a diagram that required mental rotation. The total number of questions answered correctly was the final score. The test used in the assessment of retention was identical to that used in the short-term assessment, which was developed for a previous study.8

Initial recruitment of participants for short-term assessment was by e-mail and class announcements. Participants were randomized into 2 groups via a random number generator. The groups were identified as the SWI group and the 3-D computer model (3-D) group. The medium used for laryngeal anatomy teaching was presented to the participants via computer. The content of the presentation was identical between groups except for the SWI group receiving static 2-D images in place of the 3-D images presented to the 3-D group. Participants from each group were given 45 minutes to study their respective media.

No additional anatomy instruction was given to participants before the long-term assessment. All 100 of the original participants were subsequently contacted by e-mail and invited to retake the laryngeal anatomy quiz approximately 6 months after their initial test. Students were not informed at the beginning of the study that a follow-up study would be performed 6 months later to avoid incentive for studying. Tests were returned by e-mail for grading. The original group of 100 students was identified as the short-term group. Students who retook the test were identified as the long-term group.

Statistical analysis was performed using a spreadsheet program (Microsoft Excel; Microsoft Corp, Redmond, Washington) and a statistical analysis program (SAS, version 9.2; SAS Institute, Inc, Cary, North Carolina). Measures of central tendency (mean [SD]) were first obtained for demographic data and for scores obtained in the laryngeal anatomy test. Demographic differences between the short- and long-term groups were assessed using the t test and 1-way analysis of variance for continuous variables and the χ² test for dichotomous variables. To determine whether there was a difference in test scores between the 3-D and SWI groups, unpaired t tests were performed. To determine whether there was a difference in test scores between the short- and long-term test sessions, paired t tests were performed. An a priori probability level was set at P < .05 for all the previously mentioned tests. All aspects of this study were approved by the research ethics board at the University of Western Ontario.

RESULTS

Sixty-two of the original 100 students (mean [SD] age, 23.8 [1.9] years; 56% male) retook the laryngeal anatomy test: 30 in the 3-D group and 32 in the SWI group. Detailed demographic data for the long-term group are given in Table 1. No statistically significant differences were noted between students who retested and those who did not (Table 2).

Figure 1 shows the differences in mean overall, 3-D, and factual scores for the 20-question test on laryngeal anatomy for the short- and long-term groups. The mean (SD) overall scores for the long-term group were 14.5 (2.4), 4.0 (1.3), and 10.5 (1.8), respectively, which were not found to be significantly different from those for the short-term group (P = .54, P = .07, and P = .52, respectively). The changes in mean overall, 3-D, and factual scores for the short- and long-term groups were also not found to be significant (P = .59, P = .61, and P = .71, respectively).

For the long-term retention group, the mean (SD) overall scores were 14.3 (2.2) for the 3-D group and 14.7 (2.4) for the SWI group, which were not significantly different (P = .53). The mean (SD) 3-D scores were 4.0 (1.1) for the 3-D group and 4.1 (1.5) for the SWI group, which were not significantly different (P = .70). The mean (SD) factual scores were 10.4 (1.8) for the 3-D group and 10.6 (1.9) for the SWI group, which were not significantly different (P = .58).

Figure 2A shows the overall scores for the 3-D and SWI groups comparing short- and long-term retention. For the long-term retention group, the mean (SD) overall scores were 14.3 (2.2) for the 3-D group and 14.7 (2.4) for the SWI group, which were not significantly differ-
There was no significant difference in the overall scores for the 3-D (P = .53) and SWI (P = .95) groups. There was no significant difference in the overall scores for the 3-D (P = .41) and SWI (P = .95) groups. Figure 2B shows the 3-D scores for the 3-D and SWI groups comparing long- and short-term retention. The 3-D score is the sum of the scores of the seven 3-D questions. The mean (SD) 3-D score for the 3-D group was 4.0 (1.1) vs 4.1 (1.5) for the SWI retention group, which were not significantly different (P = .70). There was a trend toward an increase in the 3-D score in both groups (P = .07 in Figure 1), with a significant increase in 3-D scores in the 3-D group only (P = .049 in Figure 2B).

There was no significant difference in mean factual scores for the 3-D group in the short-term (10.1) and long-term (10.4) groups (P = .87). No significant difference was noted in mean factual scores for the SWI group in the short-term (10.9) and long-term (10.6) groups (P = .41).

Table 2. Demographic Characteristics of Students Who Retested vs Those Who Did Not

<table>
<thead>
<tr>
<th>Variable</th>
<th>Retested (n=62)</th>
<th>Not Retested (n=38)</th>
<th>Totala (n=100)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class, No.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medicine y 1</td>
<td>20</td>
<td>15</td>
<td>35</td>
<td>.26</td>
</tr>
<tr>
<td>Medicine y 2</td>
<td>11</td>
<td>1</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Medicine y 3</td>
<td>9</td>
<td>4</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Medicine y 4</td>
<td>6</td>
<td>5</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Dentistry</td>
<td>10</td>
<td>7</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Physical therapy</td>
<td>6</td>
<td>6</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Male sex, No. (%)</td>
<td>35 (56)</td>
<td>20 (53)</td>
<td>55 (55)</td>
<td>.71</td>
</tr>
<tr>
<td>Age, mean (SD), y</td>
<td>23.7 (1.9)</td>
<td>23.8 (2.6)</td>
<td>23.8 (2.2)</td>
<td>.91</td>
</tr>
<tr>
<td>Highest education, No.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor degree, 3 y</td>
<td>12</td>
<td>5</td>
<td>17</td>
<td>.44</td>
</tr>
<tr>
<td>Honors degree, 4 y</td>
<td>46</td>
<td>28</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>Masters degree</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Science background, No. (%)</td>
<td>60 (98)</td>
<td>34 (92)</td>
<td>94 (94)</td>
<td>.15</td>
</tr>
<tr>
<td>Taken anatomy before, No. (%)</td>
<td>52 (84)</td>
<td>27 (73)</td>
<td>79 (79)</td>
<td>.19</td>
</tr>
<tr>
<td>Taken laryngeal anatomy, No. (%)</td>
<td>35 (56)</td>
<td>20 (54)</td>
<td>55 (55)</td>
<td>.82</td>
</tr>
<tr>
<td>Total score, mean (SD)</td>
<td>14.3 (2.8)</td>
<td>14.0 (2.9)</td>
<td>14.2 (2.8)</td>
<td>.59</td>
</tr>
<tr>
<td>Total score, % (SD)</td>
<td>71.7 (13.8)</td>
<td>70.19 (14.3)</td>
<td>72.0 (15.1)</td>
<td>.59</td>
</tr>
<tr>
<td>3-dimensional score, mean (SD)</td>
<td>3.7 (1.3)</td>
<td>3.8 (1.5)</td>
<td>3.72 (1.4)</td>
<td>.23</td>
</tr>
<tr>
<td>Factual score, mean (SD)</td>
<td>10.7 (2.0)</td>
<td>10.2 (1.7)</td>
<td>10.5 (1.9)</td>
<td>.59</td>
</tr>
</tbody>
</table>

a Reproduced with permission from Hu et al.1

To our knowledge, we are the first to assess the role of a 3-D educational computer model in conferring the long-term retention of anatomy instruction. Previous studies10,11 have demonstrated that in the short-term, 3-D CAI models do not confer any distinct educational advantage compared with standard didactic lecture instruction, although in most cases, they have been shown to be equally effective. In this study, we found that a low-fidelity model (SWI) is just as effective as a high-fidelity model (3-D) in teaching laryngeal anatomy and that the acquired knowledge from either educational intervention may last up to 6 months.

Psychologists have long been intrigued by the processes of learning and forgetting. In 1885, a psychologist named Hermann Ebbinghaus pioneered this field by assessing “pure learning,” that is, learning free of meaning, and studying the rate of forgetting. He used materials with little or no meaning because learning new information is influenced by context and material that the student already knows. He discovered that material is forgotten in a highly predictable, exponential manner described by what is now known as the Ebbinghaus forgetting curve.12 This curve is described by the formula $R = e^{-t/s}$, where $R$ is memory retention, $s$ is the relative strength of memory, and $t$ is time. According to Ebbinghaus’ research work, two-thirds of material is forgotten within a day. After 3 to 6 days, most students remember only 10%. After that, the curve plateaus and knowledge is stored in long-term memory. Therefore, testing 6 months after the intervention truly does test long-term retention.

There have been 2 previous studies10,11 on long-term retention in medical education, but neither was in

---

**COMMENT**

To our knowledge, we are the first to assess the role of a 3-D educational computer model in conferring the long-term retention of anatomy instruction. Previous studies have demonstrated that in the short-term, 3-D CAI models do not confer any distinct educational advantage compared with standard didactic lecture instruction, although in most cases, they have been shown to be equally effective. In this study, we found that a low-fidelity model (SWI) is just as effective as a high-fidelity model (3-D) in teaching laryngeal anatomy and that the acquired knowledge from either educational intervention may last up to 6 months.

Psychologists have long been intrigued by the processes of learning and forgetting. In 1885, a psychologist named Hermann Ebbinghaus pioneered this field by assessing “pure learning,” that is, learning free of meaning, and studying the rate of forgetting. He used materials with little or no meaning because learning new information is influenced by context and material that the student already knows. He discovered that material is forgotten in a highly predictable, exponential manner described by what is now known as the Ebbinghaus forgetting curve. This curve is described by the formula $R = e^{-t/s}$, where $R$ is memory retention, $s$ is the relative strength of memory, and $t$ is time. According to Ebbinghaus’ research work, two-thirds of material is forgotten within a day. After 3 to 6 days, most students remember only 10%. After that, the curve plateaus and knowledge is stored in long-term memory. Therefore, testing 6 months after the intervention truly does test long-term retention.

There have been 2 previous studies on long-term retention in medical education, but neither was in
anatomy. The first study was a randomized controlled study testing the ability of 3-D Web learning models to enhance traditional teaching formats for embryonic development by Marsh et al. Students were exposed to traditional lecture materials and to 3-D modules simultaneously on 2 separate occasions with a difference in short-term test scores that did not reach significance. At 16 months, the 3-D model group had a higher score than did the control group that was statistically significant. Marsh et al concluded that the 3-D modules may be more useful if used toward the later stages of learning rather than as an initial resource. These findings are consistent with observations in a previous study indicating that junior learners seem to have more difficulty in synthesizing information from complex models, such as our high-fidelity model, and may receive greater benefit when a solid foundation of information learned from lower-fidelity models is already in place. Different quizzes were used by Marsh et al to assess retention, with an initial 14-question quiz and an 8-question retention quiz using different questions. In the present study, we elected to use the same quiz for the assessment of short- and long-term retention to avoid any confounding variables regarding quiz content.

In the second study, D'Alessandro et al conducted a randomized controlled study testing the long-term instructional effectiveness of a pediatric multimedia textbook to teach common pediatric airway diseases to undergraduate medical students. In the short-term, multimedia textbooks were shown to be more effective than standard lectures and printed textbooks. The authors did not detect any significant differences between instructional methods after 11 to 22 months, with decreases in retention scores for the control and experimental groups. The present findings are similar in that the high- and low-fidelity models did not show a significant difference in test scores at 6 months, although we did not observe any significant advantage of the high-fidelity model in the short-term. An interesting finding in this study was the maintenance in knowledge over time.

In this study, there was an overall trend of increase in 3-D scores with both groups and a significant increase in the 3-D group. This may be explained by the potential of an increase in the participant’s foundation of knowledge between quizzes, which may have assisted with the synthesis of information that was initially learned. Participants in the 3-D group may have been stimulated to acquire additional knowledge based solely on interaction with the 3-D model because it has been rated to be more enjoyable than SWI. The attrition rate in the study by D’Alessandro et al may also have been a factor because it was relatively high at 50.3%, whereas the attrition rate in the present study was much lower at 38%.

As noted in a previous study, student learning styles may be an important factor in the effectiveness of CAI models when teaching anatomy in 3-D. Novice student learners may receive the most benefit from simple 2-D teaching modalities vs visually complex 3-D models simply owing to the volume of information presented at one time. The effectiveness of a modality to teach new material may increase if presented in a format that is conducive to the learner’s strengths. In addition, the role of student learning style has been shown to be important in the rate of use of teaching modalities as students with complementary styles have increased modality use. Therefore, pairing complementary teaching modalities to student learning styles may result in improved effectiveness due to optimized student exposure and presentation. Improved performances of students using SWI in a previous study may reflect the impact of learning styles because the participants were novice students whose style of learning may have been more conducive to standard, non–3-D instruction.

A factor that was not assessed in this study was the individual spatial ability of the participants. It has been shown that students have varying abilities with respect to understanding 3-D relationships and their position when manipulated. Students with a lower spatial ability underachieve in tests of anatomy deemed to be more spatially complex. As a result, students with a high level of spatial ability may receive greater benefit from a high-fidelity model.
such as the present one, which can present views that are more spatially complex than can traditional 2-D models. Furthermore, keeping in mind the disparity in student spatial ability, there may not be one solution for all learners but an increased need for instruction to be tailored to individual student abilities.

In recent times, with the increased prevalence distributed medical education and increased student enrollment, the delivery of medical school instruction of clinical skills and theory has become increasingly important. The use of model-based instruction has proved to be effective in the transfer of skills from the laboratory to the operating room, as is the case in the study by Anastakis et al.\textsuperscript{15} where skills learned on bench model were shown to transfer well to working with human models. The use of high-fidelity models, such as the computerized virtual reality bronchoscopy simulator by Chandra et al.\textsuperscript{16} and the full-scale larynx simulator used by Friedman et al.\textsuperscript{17} has been shown to be of equal effectiveness relative to low-fidelity models. The present study shows that even in the long-term, a high-fidelity model such as this is still equally as effective as a low-fidelity model. We believe that 3-D computer models are not meant to replace traditional medical instruction techniques, such as cadaver dissection and didactic lecture, but to act as valuable supplements. In the case of satellite medical sites, they may provide a practical and efficient alternative to traditional instruction that may otherwise be unavailable.

This study has several strengths. Relative to past studies, the rate of attrition after 6 months was low at 38%, resulting in a large sample size of 62. An attempt to control confounding variables was made by using consistent testing mediums and content between assessments, minimal variation between participant examination times, and randomization of participants. As a result, the only variable being tested was the influence of 3-D vs 2-D images. This study population is a homogeneous group of health care students with similar backgrounds. In addition, the 3-D model of the larynx was specifically created and designed for this study, instead of using a commercially available model. The utility of this 3-D CAI model has many practical real-world applications. As previously mentioned, distributed medical education is becoming more common in undergraduate medical education. Simulations and CAIs provide an efficient, low-maintenance, and cost-effective means of education delivery to these distant sites, as evidenced by the recent integration of this 3-D model into the University of Western Ontario first-year medical curriculum.

There are some limitations to this study. Administration of the quizzes for long-term assessment was by e-mail, and thus, the opportunity for participants to look up answers to questions or to brush up on material before taking the quiz could not be accounted for, although there would be minimal incentive or reward to do so and the randomization of participants minimizes any variation between groups. Furthermore, students were not informed at the conception of the study that a follow-up study would be conducted 6 months later. In addition, because the participant population consisted of students who were actively enrolled in medical, dental, and physical therapy programs, acquired knowledge between steps could not be controlled, although this variable should be minimal due to the randomization of participants. As with a past study, we did not extensively assess the validity of the primary outcome measure, the test on laryngeal anatomy knowledge, although initially attempts were made to confirm the construct validity. This validity testing was done by evaluation of its content by 3 experts. The study consisted of 2 groups with either SWI or 3-D instruction but did not include a control group without instruction. We believe that although this may have helped demonstrate retention of the learned information, the differences in retention between the SWI and 3-D groups being examined would remain unchanged.

In conclusion, to our knowledge, this is the first study to examine long-term retention using a 3-D CAI model. A low-fidelity model (SWI) is just as effective as a high-fidelity model (3-D instruction) in teaching laryngeal anatomy. The acquired knowledge from either educational intervention may last up to 6 months for long-term retention. A possible future direction of research is the assessment of individual student spatial ability and its effect on the effectiveness of the high-fidelity model.

Submitted for Publication: May 26, 2010; final revision received September 13, 2010; accepted December 17, 2010.

Correspondence: Kevin Fung, MD, FRCSC, Department of Otolaryngology—Head and Neck Surgery, London Health Sciences Centre, Victoria Hospital, University of Western Ontario, 800 Commissioners Rd E, London, ON N6A 5W9, Canada (kevin.fung@lhsc.on.ca).

Author Contributions: All authors had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Hu, Wilson, Ladak, Haase, and Fung. Acquisition of data: Fritz and Hu. Analysis and interpretation of data: Fritz, Hu, Wilson, and Fung. Drafting of the manuscript: Fritz, Hu, and Wilson. Critical revision of the manuscript for important intellectual content: Fritz, Wilson, Ladak, Haase, and Fung. Obtained funding: Hu and Fung. Administrative, technical, and material support: Fritz, Wilson, and Ladak. Study supervision: Fritz, Hu, Wilson, Ladak, and Fung.

Financial Disclosure: None reported.

Funding/Support: This project was supported by Research on Teaching—Small Grants Program, Teaching Support Centre, University of Western Ontario, and Faculty Support for Research on Education Grant, Schulich School of Medicine and Dentistry, University of Western Ontario.

Previous Presentation: This study was presented in part at the Annual Meeting of the Canadian Society of Otolaryngology—Head & Neck Surgery; May 24, 2010; Niagara Falls, Ontario, Canada.

REFERENCES