ASE-13-0096.R2

Research Report

Human Cadavers vs. Multimedia Simulation:

A Study of Student Learning in Anatomy

Andrew J. Saltarelli,[1] Cary J. Roseth,[2] William A. Saltarelli[3]

[1]Office of the Vice Provost for Online Learning, Stanford University, Stanford, California

[2]Department of Counseling, Educational Psychology and Special Education, College of Education, Michigan State University, East Lansing, Michigan

[3]Department of Health Promotion and Rehabilitation, College of Health Professions, Central Michigan University, Mount Pleasant, Michigan

Running title: Cadavers, multimedia simulation, and anatomy

Correspondence to: Dr. Andrew J. Saltarelli, Office of the Vice Provost for Online Learning, Stanford University, Stanford, CA 94305. USA. E-mail: ajsalts@gmail.com

ABSTRACT

Multimedia and simulation programs are increasingly being used for anatomy instruction, yet it remains unclear how learning with these technologies compares with learning with actual human cadavers. Using a multilevel, quasi-experimental-control design, this study compared the effects of *Anatomy and Physiology Revealed* (APR 2.0) with a traditional undergraduate human cadaver laboratory. APR is a model-based multimedia simulation tool that uses high-resolution pictures to construct a prosected cadaver. APR also provides animations showing the function of specific anatomical structures. Results showed that the human cadaver laboratory offered a significant advantage over the multimedia simulation program on cadaver-based measures of identification and explanatory knowledge. These findings reinforce concerns that incorporating multimedia simulation into anatomy instruction requires careful alignment between learning tasks and performance measures. Findings also imply that additional pedagogical strategies are needed to support transfer from simulated to real-world application of anatomical knowledge.

Keywords: gross anatomy education, digital anatomy, multimedia, computers in anatomical education, Interactive computer graphics, computer-aided instruction, CAI, teaching of anatomy

INTRODUCTION

One of the enduring controversies in teaching gross anatomy is whether human cadavers are uniquely beneficial to student learning (McLachlan et al., 2004; Mitchell and Stephens, 2004; Pereira et al., 2007; Ramsey-Stewart et al., 2010). Proponents argue that only human cadavers can provide the visual and tactile experiences necessary to learn and apply anatomical knowledge (Dyer and Thorndike, 2000; Aziz et al., 2002). Opponents argue that the capabilities of modern multimedia and simulation technologies outweigh the benefits, if any, of using human cadavers (McLachlan et al., 2004). This controversy has only intensified in recent years as the increasing costs of traditional anatomy laboratories must be weighed against the decreasing costs and increasing capabilities of modern technologies (McLachlan et al., 2004). In fact, medical schools in Australia (see Ramsey-Stewart et al., 2010) and the United Kingdom (see McLachlan et al., 2004) have already ceased anatomical dissection despite little empirical work examining the proposition.

Comparing APR and human cadavers in learning anatomy

The present study contributes to this debate by comparing the effects of a traditional human cadaver laboratory with a modern multimedia-based, virtual simulation learning system, *Anatomy and Physiology Revealed,* version 2.0 (APR, 2009). APR is a model-based computer simulation that constructs a real, prosected cadaver via high-resolution pictures (see Figure 1). APR also provides pre-produced computer animations that show, for example, the blood flow to the heart muscle and what happens when a blockage occurs (see Figure 2). At present, APR does not support practice-oriented simulations such as the ability to initiate and then alleviate a virtual

heart attack, nor does it support physical simulations such as providing tactile feedback to learners.

APR as a model-based simulation

One advantage of APR compared to cadaver-based instruction may be the ability to navigate efficiently through different bodily layers and structures in a three-dimensional space. Research examining model-based simulations in anatomy instruction is mixed, however, with one study reporting no significant effects on spatial and factual knowledge of anatomical structures (Keedy et al., 2011), and two other studies reporting positive effects on students' spatial understanding (Nasr, 2007; Hisley et al., 2008).

Another advantage of APR may be providing more opportunities to explore anatomical structures on their own. APR also allows students more autonomy in choosing different views, angles, and combinations of anatomical images that is not possible in traditional cadaver-based instruction. Nasr (2007) conducted a non-randomized experimental-control study and found that undergraduates who performed dissection using APR, version 1.0 (a previous version of the APR software) scored significantly better on text-based quizzes than students performing dissections on fetal pigs.

A disadvantage of APR compared to cadaver-based instruction may be the lack of haptic (sense of touch) understanding of anatomical structures. Kinesthetic learning is seen by many to be integral to anatomy instruction (Preece et al., 2013) and especially in the health professions where anatomy instruction is ultimately concerned with enhancing students' clinical practice with human patients (Dyer and Thorndike, 2000; Aziz et al., 2002). Another disadvantage of APR may be in its ability to give students too many views of anatomical structures and, so doing,

distracting them from focusing on key information. Research by Levinson and colleagues found that students given more pictorial views (beyond those considered to be "key" for learning an anatomical structure) and more control of the virtual environment performed worse than those given only "key" views of structures and less control (Levinson et al., 2007).

APR as a multimedia learning tool

In addition to providing model-based simulations, APR also provides a number of multimedia features including audio pronunciations, histological and radiological imaging, and three-dimensional animations showing the function of specific anatomical structures (see Fig. 2). APR's ability to present multimodal information to learners may be an advantage over cadaver-based instruction because, with a few clicks, students can visually identify a structure, read its label, listen to its pronunciation, and watch an animation showing its function. APR also includes a self-quiz function that may make learning more efficient than cadaver-based learning. Another advantage of APR may be its ability to present pictorial information in higher quality (accuracy) and greater quantity (different views, repeated viewings) than real cadavers which deteriorate quickly. APR also overlays clear labels of all important structures which may be more clear and accurate than those found in cadaver-based learning.

Another disadvantage of APR may be that the software's navigation and abundant options overwhelm students' cognitive processing capabilities and, as a result, impair learning (Kalyuga et al., 1999; Kirschner, 2002). If students are not able to coordinate the pictorial, audio, and text presentations found in APR, they may perform worse than if they were studying with simpler cadaver-based materials.

Present study

In summary, it remains unclear whether modern multimedia and simulation technologies may be used to replace traditional cadaver-based laboratories in anatomy instruction. The present study addresses this issue using a multilevel, quasi-experimental-control design to compare the effects of APR- and cadaver-based laboratories on undergraduate student learning. Importantly, student learning was assessed on actual (rather than digital) anatomical structures on a human cadaver. Student learning was also assessed in terms of two aspects of biomedical knowledge: identification of anatomical structures and explanation of causal mechanisms behind bodily functions (Patel et al., 2001). Thus, the present study also examines whether APR- and cadaver-based laboratories have different effects on these two outcomes.

The study's specific research questions were as follows:

1. Does student learning of anatomical knowledge differ as a function of using a multimedia and simulation technology (i.e., APR) rather than a cadaver-based laboratory?
2. If there is evidence that student learning differs between APR- and cadaver-based laboratories, are these differences conditional on students' current course grade and the instructional unit?

METHOD

Participants

This study was conducted in an undergraduate course in human anatomy at a large, public midwestern university. This semester-long course includes two lectures and two laboratories per week, all 75 minutes in length. The lecture section is taught by a full professor while laboratory sessions are taught by graduate students. Approximately 80% of students taking

the course are in their first or second year of university and pursuing baccalaureate (B.A.) degrees in exercise physiology, neuroscience, athletic training or biomedical sciences. The other 20% are in their third or fourth year and fulfilling a prerequisite requirement for professional allied health programs.

Eligibility criteria for this study included voluntary participation and signed consent of an undergraduate student between the ages of 18 and 24. Exclusion criteria included the inability to read and write in English and unwillingness to follow procedural directions. Students received extra credit for participation. In all, 77% (N = 233) of recruited students agreed to participate in the study, with n = 68 absent on the day of the study. Procedures associated with the study were reviewed and approved by the sponsoring university's institutional review board (IRB No. 151908-2).

Procedures

A quasi-experimental-control design was used, with the 14 laboratory classrooms randomly assigned to APR (7 classrooms) or cadaver-only (7 classrooms) laboratories. Laboratory classrooms randomly assigned to the cadaver-only condition in the first unit participated in the APR condition in the second unit (and vice versa), with the order of conditions counter-balanced across all 14 classrooms. Thus, all students participated in both the APR and cadaver-only laboratories. All students also worked on the same two units (cerebral spinal fluid, blood vessels), each lasting two weeks. Pre-recorded lecture screencasts (approximately 20 minutes long) were used in both the APR and cadaver-only laboratories to outline the objectives for the laboratory and to minimize differences across the seven instructors covering the 14 laboratory classrooms. All instructors were trained on the APR software.

Independent variable

The independent variable was instructional technology, with the experimental condition using APR and the control condition a prosected human cadaver. Otherwise, every effort was made to keep all instructional objectives and materials consistent between the experimental and control conditions. Specifically, at the beginning of the laboratory session, students in the cadaver-only control group were given a list of objectives and allowed free use of the five prosected cadavers for the remainder of the laboratory session (55 minutes). Students in the experimental APR group were given an identical list of objectives but then instructed to sit down at one of five laptop computers to use the APR software for the remainder of the laboratory session. Approximately four students worked at each computer, which is the same number of students that work on each cadaver. Laboratory instructors were present to answer questions.

Dependent variables

There were two quantitative dependent variables: identification and explanation. At the end of each two-week unit, students took an exam on all material covered during that unit. For this study, only those questions focusing on the laboratory session objectives were used in the analysis (see Appendix A for exam questions). Each exam contained ten open-ended questions pertaining directly to the laboratory objectives covered during laboratory sessions. Five questions on each exam were identification questions (e.g., "identify the blood vessel marked by pin number one") and were added together to comprise identification (Cronbach's $\alpha = .66$). Five other questions asked students to explain how an identified anatomical structure works (e.g.,

"name the organ that pinned blood vessel number nine supplies") and were added together to comprise explanation (Cronbach's $\alpha = .67$). Each question was worth two points.

Materials

Anatomy and Physiology Revealed 2.0 (APR, 2009) is software jointly developed by the McGraw-Hill publishing company and the University of Toledo. The software can either be installed locally on a computer or accessed via a secure website. The program is based off of high-resolution pictures of real, expertly dissected cadavers. Pictures are layered in a way that allows students to interact with and explore many dissection depths for each region of the body. At each layer, clickable "pins" reveal the names of anatomical structures with audio pronunciations, histological and radiological imaging, and three-dimensional animations showing the function of specific structures. There is also a comprehensive self-quizzing tool associated with each region of the body.

All laptop computers were 14-inch Apple iBook computers with 1.0 Ghz G4 processors and 512mb of RAM. This hardware configuration was well above the minimum requirements for APR and the software ran with very little lag when transitioning between instructional elements. The APR software was run off of CD media at each workstation.

Data Analysis

This study's data may be conceptualized as clustered repeated measures, with students (the unit of analysis) nested within classrooms, and repeated measures (Units A and B) collected on students over time. Accordingly, we used a series of linear mixed models (LMMs; Fitzmaurice et al., 2007) to compare the relative effectiveness of APR and cadaver conditions

while accounting for the clustered structure of the data and the correlated error terms associated with the repeated measures. LMMs can be viewed as multilevel, or hierarchical linear models with individual- and cluster-level equations (Raudenbush and Bryk, 2002). For example, this study's data can be considered to have three levels, with Level 3 representing the classrooms, Level 2 the students, and Level 1 the repeated measures. Appendix B includes additional information about LMMs and the specific unconditional and conditional models used for the analysis.

We used the "top-down" strategy for model building (West et al., 2007), starting with a model that included fixed effects for all of our covariates (including interactions between the covariates), then selecting a covariance structure for the residuals, and finally using statistical tests to determine whether some fixed effects may be removed from the model. To estimate parameters and test statistics, we used the restricted maximum likelihood method (REML), the Kenward and Roger (1997) method for degrees of freedom, and a significance level, $\alpha = .05$, in the PROC MIXED procedure of SAS 9.2. (SAS Institute Inc., Cary, NC). Selecting an appropriate variance-covariance structure for the residuals was based on the procedures outlined by Fitzmaurice et al. (2004) and West et al. (2007), with the comparison (not presented) of deviance and fit statistics (e.g., -2LL) indicating best fit. Throughout the model-building process, normal probability plots were used to screen residuals for violations of assumptions (e.g., normality and constant variance) and outliers.

RESULTS

There were 14 laboratory classrooms, with an average of 15.28 students per classroom and a range of 11 to 21 students per section. Across the two laboratory units, 3 of 108 students in

the APR condition had missing data in Unit A, and 4 of 109 students in the cadaver condition had missing data in Unit B, resulting in only 7 missing out of 434 total data points. Visual inspection of conditional residual plots supported the statistical validity of results. Table 1 shows the descriptive information for the variables across the two units and experimental conditions. Table 2 shows inter-correlations among the two measures of anatomical knowledge and current course grade.

For student learning of anatomical knowledge, two types of analyses were performed: (1) unconditional analyses that examined how much of the variation in learning was due to student- and classroom-level factors, and (2) conditional analyses that compared the relative effects of APR and cadaver laboratories while accounting for student characteristics that may be confounded with learning. We detail the results of the conditional analyses here, as the study's primary goal was to compare the APR and cadaver conditions. Table 3 reports the results for the unconditional and conditional models for both learning outcomes and the key parameter, β_3, is bold-faced.

Controlling for current course grade and differences associated with the laboratory units, the slope term was negatively statistically significant for both Identification ($\beta_3 = -1.24$) and Explanation ($\beta_3 = -0.92$). As displayed in Figure 1, this indicates that the APR condition offered a significant disadvantage over the cadaver condition for both learning outcomes. For Identification, this disadvantage corresponded to a 19 and 13.6% change for Units A and B, respectively, and for Explanation a 15.7 and 8.7% change.

Interestingly, there were no statistically significant interactions with instructional condition, suggesting that the effect of the cadaver and APR conditions did not depend on the Unit or current course grade. Still, both Unit and Grade did account for significant variation in

student learning. For instructional Unit, the slope term was negatively statistically significant for both Identification (β_1 = -4.18) and Explanation (β_1 = -5.58), even controlling for variation associated with current course grade and the effect of the cadaver and APR conditions. This indicates that student learning was lower in Unit B compared to Unit A. This decrease varied as a function of current course grade, however, as indicated by the positively statistically significant interaction term, β_4. This indicates that student learning in Unit B did not decrease as significantly among students with higher course grades.

Looking at the random effects, we calculated the proportion of variance explained by adding the student- (Grade) and classroom-level (Condition) predictors. For Identification, the proportion of between-student variance explained by the conditional model was (1.98 – 0.99) / 1.98 = 0.50, or 50%, and for Explanation the proportion of between-student variance explained was 27%. Despite these reductions, both the residuals and student-level variance remained statistically significant for both measures, indicating that unexplained variation remained. In contrast, the classroom-level variance estimates were not statistically significant for any of the models. This indicates that classroom-level mean student learning scores were homogenous after controlling for variation associated with students and repeated measures.

DISCUSSION

Multimedia and simulation programs increasingly are being used for anatomy instruction. Little is known about how learning with these technologies compares to learning with actual human cadavers, especially when learning is assessed on actual (rather than digital) anatomical structures on a human cadaver. This study addressed this issue by using a multilevel, quasi-experimental-control design to compare the effects of Anatomy and Physiology Revealed 2.0

(APR) and human cadavers on students' learning of two forms of anatomical knowledge: identification and explanation.

Results show that the APR condition offered a significant disadvantage over the cadaver condition for both learning outcomes, even after controlling for students' current course grade, differences associated with the instructional units, and the nested quality of student scores in different laboratory classrooms. These findings suggest that learning with APR and human cadavers results in different outcomes, at least as measured by identification and explanation of actual anatomical structures on a human cadaver. This caveat is important as this study's results contrast with other studies reporting the positive effects of simulations on student learning as measured by digital images of anatomical structures (Hisely et al., 2007; Nasr, 2007).

A likely explanation of these different findings involves transfer of learning, broadly defined as the extent to which knowledge learned in one context is applied in another (for a review see: Barnett and Ceci, 2002). Transfer of learning is at the center of the controversy in anatomy instruction to the extent that the debate is ultimately concerned with whether students can apply knowledge learned with a multimedia simulation technology (one context) on actual human beings (a second context). From a transfer perspective, for example, the Hisely et al. (2008) and Nasr (2007) studies' results suggest that learning with human cadavers and fetal pigs were disadvantaged to the extent that they were unable to transfer their knowledge to digital images of anatomical structures. Likewise, in the present study, students learning with APR were disadvantaged to the extent they were unable to transfer their knowledge to actual human cadavers. Thus, even as this and prior simulation studies have not examined transfer in the traditional sense of addressing questions about identical elements (e.g., Thorndike, 1913), general principles (e.g., Judd, 1908), or preparation for future learning (e.g., Bransford and

Schwartz, 1999), the pattern of findings highlights the importance of understanding the nature and extent of transfer in learning human anatomy with different instructional technologies.

More specifically, this study's findings highlight the need for future research to determine how and when to integrate multimedia simulation experiences with actual cadaver-based experiences, as many anatomy instructors can and will supplement traditional laboratories with various technological alternatives (Collins, 2008). This type of research would be especially relevant to the health professions, where the cost of maintaining traditional laboratories will likely continue to be a problem (James et al., 2004) and the value of technological alternatives to human cadavers must ultimately be measured by students' application of knowledge on actual human patients in real-world clinical contexts.

LIMITATIONS

This study's results are limited by the length of the study, characteristics of the sample, the type of instructional technology (i.e., APR), and reliability of the measures used. Specifically, it remains unclear whether student learning would improve given more time to use APR and become more efficient and comfortable using the technology. A related limitation is that this study did not control for students' perceptions of APR, as the perceived usefulness and ease of using APR may have contributed to their motivation to use the technology (Venkatesh, 2000). Future research should control for users' perceptions of technology and explore its relationship to transfer of learning.

Regarding the sample, the extent to which the study's results reflect sampling error also remains unclear. Indeed, this study's results contrast with previous research by Hisley and

colleagues who found no significant differences between physical and digital dissection on medical students' percentages of correct responses (Hisley et al., 2008). Future research is needed to determine whether these different findings are due to differences in the samples (undergraduates versus medical students), the simulation technologies (APR versus three-dimensional volume modeling of whole-body CT and MRI image sets), exploration method (prosection versus dissection) or the type of assessment (e.g., identification and explanation of anatomical structures on human cadavers versus identification and spatial ordering of digital models).

Finally, it should also be noted that the reliability of the five-item identification and explanation measures was not ideal ($\alpha = 0.66$ and 0.67, respectively), despite the fact that both measures correlated positively with each other and with students' current course grade. Future research should use measures of student learning with larger numbers of items.

CONCLUSIONS

This study contributes to the literature in several important ways. First, this study's findings have strong internal validity, as the use of random assignment, a quasi-experimental-control design, and standardized instructional methods all served to strengthen confidence that differences between conditions were due to the independent variable. Second, by using modern multilevel data analyses to account for classroom-, student-, and content-level influences, this study provides a robust view of the varied sources of influence on learning anatomy in authentic classroom contexts. Third, and perhaps most importantly, this is the only known study to compare multimedia simulation and cadaver-based laboratories on human cadaver-based measures of anatomical structures. This provides a baseline for future research concerned with

how and when to supplement traditional anatomy instruction with multimedia simulation technologies. This also provides a critical data point in the debate over fully replacing human cadavers with modern technologies such as APR. Clearly, the effective use of learning technologies involves more than replacing old technologies with new. The effective use of learning technologies such as APR will require that anatomy instructors simultaneously consider technology, pedagogy, and content, and how changes in one inevitably affects the others (Mishra and Koehler, 2006).

NOTES ON CONTRIBUTORS

ANDREW J. SALTARELLI, Ph.D., is an Instructional Designer in the Office of the Vice Provost for Online Learning at Stanford University, Stanford, California. In both his research and instructional design activities, he seeks to better understand the intersection of social psychology and instructional technology.

CARY J. ROSETH, Ph.D., is an associate professor in the Department of Counseling, Educational Psychology and Special Education, College of Education at Michigan State University, East Lansing, Michigan. He is also Director of the Hybrid Ph.D. Program in Educational Technology. His research examines how the presence of others - and especially peers - affects educational outcomes.

WILLIAM A. SALTARELLI, Ph.D., is a professor and Director of Human Anatomy in the Department of Health Promotion and Rehabilitation, College of Health Professions at Central Michigan University, Mount Pleasant, Michigan. His research interests include cardiovascular disease risk factors in children, pediatric responses to exercise, and anatomy instruction.

LITERATURE CITED

APR. 2009. Anatomy and Physiology Revealed: An Interactive Cadaver Dissection Experience. CD-ROM, version 2.0. Columbus, OH: McGraw-Hill.

Aziz MA, Mckenzie JC, Wilson JS, Cowie RJ, Ayeni SA, Dunn BK. 2002. The human cadaver in the age of biomedical informatics. Anat Rec 269:20–32.

Barnett SM, Ceci SJ. 2002. When and where do we apply what we learn? A taxonomy for far transfer. Psychol. Bull 128:612–637.

Bransford JD, Schwartz DL. 1999. Rethinking transfer: A simple proposal with multiple implications. Rev Res Educ 24:61–100.

Clark RE. 1994. Media will never influence learning. Educ Tech Res Dev 42:21–30.

Collins JP. 2008. Modern approaches to teaching and learning anatomy. Brit Med J 337:665–667.

Dyer GS, Thorndike ME. 2000. Quidne mortui vivos docent? The evolving purpose of human dissection in medical education. Acad Med 75:969–979.

Fitzmaurice GM, Laird NM, Ware JH. 2004. Applied Longitudinal Analysis. 1st Ed. Hoboken, NJ: John Wiley & Sons, Inc. 536 p.

Hisley KC, Anderson LD, Smith SE, Kavic SM, Tracy JK. 2008. Coupled physical and digital cadaver dissection followed by a visual test protocol provides insights into the nature of anatomical knowledge and its evaluation. Anat Sci Educ 1:27–40.

Judd CH. 1908. The relation of special training to general intelligence. Educ Rev 36:28–42.

Kalyuga S, Chandler P, Sweller J. 1999. Managing split-attention and redundancy in multimedia instruction. Appl Cognit Psychol 13:351–371.

Keedy AW, Durack JC, Sandhu P, Chen EM, O'Sullivan PS, Breiman RS. 2011. Comparison of traditional methods with 3D computer models in the instruction of hepatobiliary anatomy. Anat Sci Educ 4:84–91.

Kenward MG, Roger JH. 1997. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 53:983–997.

Kirschner PA. 2002. Cognitive load theory: Implications of cognitive load theory on the design of learning. Learn Instr 12:1–10.

Levinson AJ, Weaver B, Garside S, McGinn H, Norman GR. 2007. Virtual reality and brain anatomy: A randomised trial of e-learning instructional designs. Med Educ 41:495–501.

Mayer RE, Moreno R. 2002. Aids to computer-based multimedia learning. Learn Instr 12:107–119.

McLachlan JC, Bligh J, Bradley P, Searle J. 2004. Teaching anatomy without cadavers. Med Educ 38:418–424.

Mitchell BS, Stephens CR. 2004. Teaching anatomy as a multimedia experience. Med Educ 38:911–912.

Mishra P, Koehler MJ 2006. Technological pedagogical content knowledge: A framework for teacher knowledge. Teach Coll Rec 108:1017–1054.

Nasr P. 2007. Impact of multimedia technology on academic performance and student perception in the anatomy laboratory. Ohio Assoc Two Year Coll J 31:30–36.

Parker LM. 2002a. Anatomical dissection: Why are we cutting it out? Dissection in undergraduate teaching. ANZ J Surg 72:910–912.

Parker LM. 2002b. What's wrong with the dead body? Use of the human cadaver in medical education. Med J Aust 176:74–76.

Patel VL, Arocha JF, Kaufman DR. 2001. A primer on aspects of cognition for medical informatics. J Am Med Inform Assoc 8:324–343.

Pereira JA, Pleguezuelos E, Merí A, Molina-Ros A, Molina-Tomás MC, Masdeu C. 2007. Effectiveness of using blended learning strategies for teaching and learning human anatomy. Med Educ 41:189–195.

Preece D, Williams SB, Lam R, Weller R. 2013. "Let's Get Physical": Advantages of a physical model over 3D computer models and textbooks in learning imaging anatomy. Anat Sci Educ 6:216–224.

Ramsey-Stewart G, Burgess AW, Hill DA. 2010. Back to the future: Teaching anatomy by whole-body dissection. Med J Australia 193:668–671.

Raudenbush SW, Bryk AS. 2002. Hierarchical Linear Models: Applications and Data Analysis Methods. 2nd Ed. Newbury Park, CA: Sage Publications, Inc. 491 p.

Thorndike EL. 1913. An Introduction to the Theory of Mental and Social Measurements. 2nd Ed. New York, NY: Teachers College Press. 212 p.

Venkatesh V. 2000. Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. Inform Syst Res 11:342–365.

West BT, Welch KB, Galecki AT. 2007. Linear Mixed Models: A Practical Guide Using Statistical Software. 1st Ed. Boca Raton, FL: Chapman and Hall / CRC Press. 376 p.

APPENDIX A

Identification and Explanatory Examination Questions

Examination A (Blood Vessels):

1. Identify the blood vessel

2. Identify the blood vessel that #1 originated from

3. Identify the blood vessel

4. Identify the blood vessel that #3 originated from

5. Identify the blood vessel

6. Identify the blood vessel that #5 flows into

7. Identify the blood vessel

8. Name the structure that #7 drains

9. Identify the blood vessel

10. Name the organ that #9 supplies

Examination B (Cerebral Spinal Fluid):

1. Identify the structure

2. What structure would CSF flow into next after #1?

3. Identify the space

4. What space would come before #3 in the flow of CSF?

5. Identify the structure

6. What space would come before #3 in the flow of CSF?

7. Identify the structure

8. What structure would come before #3 in the flow of CSF?

9. Identify the structure

10. What structure would CSF flow into next after #9?

APPENDIX B

Linear mixed models (LMMs) were used to compare the relative effectiveness of the APR and cadaver conditions while accounting for the clustered structure of the data and the correlated error terms associated with the repeated measures.

Unconditional Linear Mixed Models

We used the following unconditional (i.e., no covariates) LMM for each variable to examine how much of the variation in student learning was due to student- and classroom-level factors:

$$Y_{tij} = \beta_0 + \beta_1 \mathrm{Unit}_t + \mu_{0j} + \mu_{0i|j} + \varepsilon_{tij}, \qquad (A.1)$$

where Y_{tij} is the learning score at time t (t = 1, 2, corresponding to Unit A and B) of the ith student nested within classroom j. By using the values 0 and 1 for Unit_t, the intercept, β_0, is the grand mean learning score on Unit A and β_1 is the difference between Unit A and Unit B grand means. The terms, u_{0j} and $u_{0i|j}$, represent the random classroom- and student-level effects associated with the intercept, respectively, and the term, ε_{ij} is the level-one residual.

Conditional Linear Mixed Models

To compare the relative effectiveness of APR and cadaver conditions, the following conditional LMM was used for each variable:

$$Y_{tij} = \beta_0 + \beta_1 \mathrm{Unit}_t + \beta_2 \mathrm{Grade}_{ij} + \beta_3 \mathrm{Cond}_j + \beta_4 \mathrm{Unit}_t \mathrm{Grade}_{ij} + \mu_{0j} + \mu_{0i|j} + \varepsilon_{tij}, \qquad (A.2)$$

where Y_{tij} is the learning score at the t-th instructional Unit (t = 1, 2, corresponding to Unit A and B) of the ith student nested within classroom j. The parameters β_0 through β_4 represent the fixed effects associated with the intercept, Unit B effect (coded 0 and 1), student-level course grade, and classroom-level Condition (cadaver coded 0 and APR coded 1). The terms, u_{0j} and $u_{0i|j}$,

represent the random classroom- and student-level effects associated with the intercept, respectively, and ε_{ij} is the level-one residual. The distributions of the random effects associated with the classroom- and student-specific intercepts are $u_{0j} \sim N(0,\ \sigma^2_{\text{int: classroom}})$ and $u_{0i|j} \sim N(0,\ \sigma^2_{\text{int: student(classroom)}})$, respectively. And the distribution of the residuals, ε_{ij}, associated with repeated measures on the same student is $\begin{pmatrix}\varepsilon_{1ij}\\ \varepsilon_{2ij}\end{pmatrix} \sim N(0, R_{ij})$. Guided by likelihood ratio (i.e., -2LL) tests, R_{ij} was specified as $\sigma^2 I_2$ for the Identification model and, for the Explanation model, as a banded main diagonal (i.e., type = un(1)) where the residuals are uncorrelated and the residual variances differ at different time points (i.e., for Unit A and B).

Hierarchical Model Specification

The model defined in Equation A.2 can also be specified in hierarchical (HLM) terms, with three levels corresponding to the contributions of the classrooms, students, and repeated measures to the variability in student learning scores.

The Level 1 (Repeated Measures) model is:

$$Y_{tij} = \beta_{0i|j} + \beta_{1i|j}\text{Unit}_t + \varepsilon_{tij}. \tag{A.3}$$

This model implies that the learning score (Y_{tij}) of the ith student nested within classroom j changes as a function of the t-th instructional Unit (t = 1, 2, corresponding to Unit A and B). This model also implies that student-specific intercepts ($\beta_{0i|j}$) and slopes ($\beta_{1i|j}$) depend on other fixed and random effects in the Level 2 model shown below.

The Level 2 (Student) model is:

$$\beta_{0i|j} = \beta_{0j} + \beta_2 \text{Grade}_{ij} + u_{0i|j} \tag{A.4}$$

$$\beta_{1i|j} = \beta_{1j} + \beta_4 \text{Grade}_{ij}.$$

This model implies that the intercepts, $\beta_{0i|j}$, for student i nested within classroom j depends on the intercept specific to the jth classroom, β_{0j}, and the student-specific covariate (Grade), and a random effect associated with the student, $\mu_{0i|j}$. The student-specific slope for Unit, $\beta_{1i|j}$, depends on the student-specific time effect, β_{1j}, and the student-specific covariate (Grade). We note that we treat the student-specific slope for Unit as a fixed effect (i.e., we do not include a random effect), assuming that differences between Units A and B are the same across students. This assumption is appropriate given that student-variation in Unit effects are not part of the research questions (Raudenbush and Bryk, 2002).

Finally, the Level 3 (Classroom) model is:

$$\beta_{0j} = \beta_0 + \beta_3 \mathrm{Condition}_j + u_{0j} \qquad \text{(A.4)}$$

$$\beta_{1j} = \beta_1.$$

This model implies that the classroom-specific intercepts, β_{0j}, depends on the overall fixed intercept, β_0, and the single covariate measured at the classroom level, Condition (cadaver coded 0, APR coded 1), and a random effect associated with the classroom, μ_{0j}. The classroom-specific slope for Unit, β_{1j}, is equivalent to the overall fixed Unit effect, β_1. As in Level 2, we treated the classroom-specific slope for Unit as a fixed effect because classroom-variation in Unit effects was not part of the research questions.

FIGURE LEGENDS

Figure 1. Anatomy and Physiology Revealed (APR) software's simulation features, as illustrated by a lateral view of the cardiovascular system in the head and neck. The left-hand figure (A) shows the Level 1 subcutaneous view, while the right-hand figure (B) shows the Level 4 subcutaneous view.

Figure 2. Anatomy and Physiology Revealed (APR) software's multimedia features, as illustrated by (A) a labeled-animation showing blood flow in the brain and (B) radiographic image of carotid arteries, and (C) histologic image of cells comprising the wall of the aorta.

Figure 3. Effects of cadaver-only and Anatomy and Physiology Revealed (APR) on identification and explanation measures of student learning (out of 10 possible points). For both measures, results showed that student learning was greater under the cadaver-only condition compared to APR and there was no evidence that these effects varied as a function of instructional Unit, though student learning was statistically significantly lower in Unit B compared to Unit A.

Table 1

Descriptive Data

Level 3 – Laboratory classrooms						
		j	M	±SD	Min	Max
Age		14	20.34	±0.73	19.27	22.00
Grade		14	0.82	±0.03	0.77	0.89
Level 2 – Students						
		i	M	±SD	Min	Max
Age	Male	89	20.44	±1.82	18	29
	Female	125	20.31	±2.23	18	40
Grade	Male	89	0.81	±0.11	0.43	0.99
	Female	125	0.83	±0.10	0.50	1.00

Level 1 – Repeated measures across Units A and B											
		Cadaver					APR software				
Measure	Unit	n	M	±SD	Skewness	Kurtosis	n	M	±SD	Skewness	Kurtosis
Identification	A	109	7.93	±2.10	-1.04	0.78	105	6.42	±2.61	-0.29	-0.96
	B	105	7.04	±2.56	-0.52	-0.67	108	6.08	±2.89	-0.39	-0.82
Explanation	A	109	7.50	±2.02	-0.60	-0.04	105	6.32	±2.25	-0.39	-0.24
	B	105	6.18	±2.98	-0.23	-1.10	109	5.64	±2.85	-0.24	-0.65

Note. Grade is the student course grade (range = 0 to 1.0) at time of study; j refers to classrooms; i refers to students; n refers to assessment; M = mean; SD = standard deviation; Min = minimum; Max = maximum.

Table 2

Inter-correlations among Measures of Anatomical Knowledge and Course Grades

Measures	Unit	2	3	4	5
1. Grade		0.39	0.52	0.31	0.44
2. Identification	A		0.32	0.53	0.36
3. Identification	B			0.28	0.78
4. Explanation	A				0.34
5. Explanation	B				

Note. All correlations statistically significant at $p < 0.01$.

Table 3

REML Parameter Estimates for Three-Level Model of Student Learning

	Identification		Explanation	
Parameters	Unconditional	Conditional	Unconditional	Conditional
Fixed effects				
Intercept, β_0	7.18[b] (0.22)	-0.12 (1.23)	6.92[b] (0.21)	1.77 (1.10)
Unit, β_1	-0.64[d] (0.21)	-4.18[c] (1.52)	-1.02[b] (0.20)	-5.58[b] (1.54)
Grade, β_2		9.55[b] (1.47)		6.77[b] (1.32)
Condition, β_3		**-1.24[b] (0.19)**		**-0.92[b] (0.18)**
Unit X Grade, β_4		4.28[d] (1.82)		5.52[c] (1.85)
Random effects				
Residual[a]	6.64[b] (0.46)	3.87[b] (0.37)	4.53[b] (0.43)	2.61[b] (0.43)
Residual, $\sigma^2_{\text{Unit B}}$				5.43[b] (0.63)
Student-level variance	1.98[b] (0.49)	0.99[c] (0.34)	2.03[b] (0.48)	1.48[b] (0.37)
Classroom-level variance	0.23 (0.20)	0.22 (0.16)	0.19 (0.18)	0.00 (-)
Fit statistics				
-2LL	2013.9	1882.8	2006.2	1908.1
AIC	2019.9	1888.8	2012.2	1914.1
BIC	2021.8	1890.7	2014.1	1916.0

Note. The intercept, β_0, is the grand mean learning score on Unit A and β_1 is the difference between Unit A (coded 0) and Unit B (coded 1) grand means; grade is the student course grade (range = 0 to 1.0) at time of study; β_3 is the difference between the cadaver (coded 0) and APR (coded 1) conditions; standard errors (SE) follow parameter estimates in parentheses; -2LL = -2 log likelihood; AIC = Akaike's Information Criterion; BIC = Bayes Information Criterion.

[a] For all unconditional models and Identification conditional model, residual variance (i.e., the $\mathbf{R}_{ij}$ matrix) was specified $\sigma^2\mathbf{I}_2$. For the Explanation conditional model, residual variance was specified as a banded main diagonal where the residuals are uncorrelated and the residual variances differ at different time points.

[b] $p < 0.001$, [c] $p < 0.01$, [d] $p < 0.05$