Human clay models versus cat dissection: how the similarity between the classroom and the exam affects student performance

John R. Waters,1 Peggy Van Meter,2 William Perrotti,3 Salvatore Drogo,3 and Richard J. Cyr1

Departments of 1Biology and 2Educational Psychology, The Pennsylvania State University, University Park, Pennsylvania; and 3Life Science Department, Mohawk Valley Community College, Utica, New York

Submitted 20 April 2009; ; accepted in final form 7 December 2010

Waters JR, Van Meter P, Perrotti W, Drogo S, Cyr RJ. Human clay models versus cat dissection: how the similarity between the classroom and the exam affects student performance. Adv Physiol Educ 35: 227–236, 2011; doi:10.1152/advan.00030.2009.—This study examined the effect of different anatomic representations on student learning in a human anatomy class studying the muscular system. Specifically, we examined the efficacy of using dissected cats (with and without handouts) compared with clay sculpting of human structures. Ten undergraduate laboratory sections were assigned to three treatment groups: cat dissection only, cat dissection with handouts, and human clay sculpting with handouts. Exams included higher-order questions that presented novel anatomic images and scenarios that the students did not practice in class. The higher-order anatomy exam questions varied the degree to which students in the different treatments had to transform the anatomic representation studied during laboratory activities to match the representation used in the exam questions. In this respect, exam questions manipulated the similarity between the surface features of the anatomic representations used in the classroom versus the exam. When identifying anatomic structures presented in a photograph or diagram, student performance improved significantly when transformation demands decreased, i.e., students in the human clay sculpting treatment group performed best on human anatomy questions and students in the cat dissection treatment group performed better on cat anatomy questions (independent of the use of handouts). There were similar, but nonsignificant, trends when students were asked functional anatomy questions presented in human and cat contexts. On survey questions designed to measure student attitudes about dissection versus nonanimal alternatives, students typically preferred the method used in their treatment group, suggesting that student preference is too fluid to factor into curricular decisions. When designing curricula, instructors must choose anatomic representations that support their course goals. Human representations are most effective when teaching the human muscular system.

A DISSECTION EXPERIENCE has been part of the study of human anatomy for centuries and has traditionally focused on the dissection of a human cadaver (8). However, many institutions do not have the resources to support a human cadaver teaching laboratory and rely instead on having students dissect animals, such as cats, when studying human anatomy. While there is limited research comparing invertebrate (4, 6), frog (7, 10, 22), rat (3, 20), and fetal pig (11) dissections to other alternatives in high school and undergraduate introductory biology classrooms, there is only one published study (26) that investigated the efficacy of a cat dissection in an introductory human anatomy course using novel higher-order anatomy questions at the undergraduate level.

In 2005, Waters et al. (26) reported that students performing cat dissections in an introductory human anatomy laboratory earned significantly poorer test scores than students who sculpted human anatomic structures from clay. These treatment differences may have been due to an inherent superiority of human clay sculpting compared with cat dissection. It is possible, for example, that the clay sculptures are a much simpler representation of anatomy than a dissected cat specimen. When using a cat specimen to study the anatomy of the muscular system, a muscle may be surrounded by connective tissue, blood vessels, nerves, or other organs, and together, these structures form a much more complex (albeit realistic) anatomic representation. In contrast, even though the clay sculpture is a much simpler representation, this simplicity may promote learning. This explanation of the Waters et al. (26) results is consistent with research demonstrating that students studying a simple diagram of cardiac blood flow earned higher test scores than students who studied a more realistic, but more complex, diagram (1). However, in an unpublished study comparing a human clay sculpting experience to a proscribed cadaver lesson (see ACKNOWLEDGEMENTS), there were no significant differences in student performance as measured by a muscular system exam, suggesting that the increased complexity of a biological teaching specimen does not always interfere with learning muscular system anatomy.

The advantage that Waters et al. (26) found for human clay sculpting over cat dissection could also be attributed to the constructive processes of sculpting. Students in the clay sculpting treatment had to build anatomic representations by sculpting each individual muscle and placing these on a plastic human skeleton. In contrast, students in the cat dissection study used preconstructed anatomic structures in a provided, intact representation. The difference between human clay sculpting and cat dissection then may mirror the differences found between studying constructed and provided representations in other research studies on science learning (16, 24, 25). Students in Waters et al.’s (26) cat dissection treatment used a laboratory manual to guide their dissections. Although students in the human clay sculpting group did not have access to the laboratory manual, this manual was supplemented with a handout that explained the order to sculpt each muscle as well as specific questions directing students to consider the action of each muscle. Thus, a reasonable hypothesis to explain the pattern of results reported by Waters et al. (26) is that prompts within the supplemental handout may have given students in the human clay sculpting treatment an advantage over their classmates in the dissection treatment.

Address for reprint requests and other correspondence: J. R. Waters, Dept. of Biology, The Pennsylvania State Univ., University Park, PA 16802 (e-mail: johnwaters@psu.edu).
The present study tested this hypothesis by including two cat dissection treatment groups: one that used the same laboratory manual to guide their dissection and a second that used the laboratory manual supplemented with a handout that listed specific muscles and included questions about the action of each muscle. If the presence of a supplemental handout has no effect on student performance, then we expect there to be no detectable differences between the cat dissection with no handout group versus the cat dissection plus a handout group.

A second alternative explanation for the Waters et al. (26) results lies in how human anatomy concepts were represented in the classroom versus on the exams. Human cadavers, anatomy software, anatomic models, and cat dissections are all representations of anatomy that can be used to learn anatomic structures. Although the conceptual structure of the anatomy content is comparable across these representations, the surface features may vary. A question on an anatomy exam may use an anatomic representation in which the surface features are identical or closely match the representation students used during classroom study or the question may reference very different representations. When the surface features of the anatomic representations used on the exam are very different from those of the representation used during study, a student must transform his/her knowledge of the studied representation to match that of the tested representation. Specifically, when a student studies cat dissections but is asked human anatomy questions, the student must transform his/her knowledge of cat anatomy to match a human representation.

All of the higher-order muscular system questions used in the Waters et al. study (26) used a human anatomic representation to test human anatomy concepts. Therefore, the overlap between the surface features of studied and tested representations was closer for students in the human clay sculpting treatment than for the students in the cat dissection treatment (26). It is possible then that the higher exam scores obtained by students in the human clay sculpting treatment may have been due to the match between the studied and tested representations. Students in the cat dissection treatment may have been at a significant disadvantage on the exam because these students had to transform their knowledge representation of the studied structure to match the test questions to a greater degree than did the students in the human clay sculpting treatment.

We tested this hypothesis by using four different representations in our posttest exam questions: a cat, a human, a horse, and a frog. These representations vary the degree to which students in the different treatments had to transform their studied representation to match the tested representation. Specifically, students in the human clay sculpting treatment would have to transform their studied representation while answering questions on cat anatomy, but these questions would be a close match for students in the cat dissection treatments. Conversely, the human anatomy questions were a close match for students in the human clay sculpting treatment but require students in the cat dissection treatments to transform their studied representation. No treatment groups studied the anatomy of the horse muscular system, but the surface features of this representation more closely matched those of the cat because both are mammalian quadrupeds. Consequently, we expected the horse anatomy questions to be a close match for students in the two cat dissection treatments. If the results of the Waters et al. (26) study can be attributed to differences in the surface feature match between studied and tested representations, then we expected that each treatment group would perform better on exam questions with a close surface feature match compared with the treatment group(s) for whom the questions required the studied representation to be transformed. We expected no treatment effects on student performance for frog anatomy questions, since the dissimilarity between amphibians and mammals means that there was a poor match between the surface features of studied and tested representations for all treatments. On the other hand, if human clay sculpting is an inherently better learning method than cat dissection, we would expect to find that students in the human clay sculpting treatment score higher on these frog anatomy questions compared with students in either of the cat dissection treatments. In summary, we tested whether or not the results reported by Waters et al. (26) may be due to a handout effect or perhaps to the degree of mental transformation necessary to apply cat anatomy to human anatomy when answering higher-order questions.

To test if there was a treatment effect on student attitudes, students were also surveyed about their preferences to study anatomy via dissection versus models, about the value of a dissection experience when one has to learn both the name and function of a structure, and about their general attitude toward anatomy courses. We expected there to be no differences before the beginning of the experiment, but student attitudes may change as they have the chance to work with different types of specimens.

METHODS

Course Description

This study was approved by the Institutional Review Board of The Pennsylvania State University and was based on the performance of 222 students enrolled in 10 laboratory sections of an introductory human anatomy course taught at the University Park campus of The Pennsylvania State University during the spring 2005 semester. The course has no prerequisites, and most of the students were majoring in Nursing, Kinesiology, or another allied health field. Almost all of the students were of typical college age (18–25 yr old). Students enrolled in this course all attended a single combined lecture section that met 2 times/wk for 50 min each meeting and also attended a single laboratory section that met 2 times/wk for 115 min each meeting. Individual laboratory sections enrolled between 20 and 24 students/section and were taught by an undergraduate or a graduate teaching assistant. All teaching assistants were supervised by, and met weekly with, the course coordinator (J. R. Waters).

Experimental Design

The lecture portion of the course was not manipulated and remained the same for all students during the entire semester. This study used a quasiexperimental design to compare the effectiveness of three types of instructional materials used during a unit on the muscular system. The first two treatments are a replication of the clay sculpting and cat dissection experiences that were tested in Waters et al. (26). In the third treatment, the students performed a cat dissection but also received a handout that was parallel to the handout used in the clay sculpting treatment. The effects of these instructional materials were evaluated by performance on a laboratory exam. Although this exam was part of the students’ coursework, items were designed to inform the questions of this study. To avoid an instructional bias, the laboratory teaching assistants who led the laboratory sections were not shown the specific exam questions until after each instructional unit was completed.
In the periods of the semester before and after the experimental phase of the course, all students experienced identical laboratory materials and activities. Before the experimental period, students studied the skeletal system; the nervous and urogenital systems came after the experimental phase. Student performance on these exams was used to ensure that differences across treatments could not be attributed to any factor other than experimental differences.

**Instructional Treatments**

When studying the anatomy of the muscular system, laboratory sections were assigned to one of the three treatment groups. In the traditional group, students performed a cat dissection that was primarily guided by photographs and diagrams of both cat and human structures in their laboratory manual. In the supplemented dissection group, students also performed a cat dissection guided by the same laboratory manual but were also provided a 19-page supplemental handout that emphasized a hands-on investigation of the structure and function of the muscles. The supplemental handout guided the order of the students’ dissection and, for each muscle, required students to first observe its size and shape, then palpitate the attachment points, and finally to pull on the muscle and deduce its action.

In the human clay sculpting treatment, students studied the structure and function of the same muscles using the same laboratory manual but instead of a cat dissection experience, students in this treatment sculpted human muscles from clay onto an 18-in.-tall plastic human skeleton model purchased from Zahourek Systems (Loveland, CO). This treatment also included a supplemental handout. The handout used in this treatment was parallel to the one used by the supplemented cat dissection group and emphasized an investigative approach to the functional anatomy of the muscular system. The handout listed the order to build the muscles onto the skeleton model, asked the students to record the attachment points, and then directed the students to deduce the muscle’s actions by attaching a piece of string to the muscle’s attachment points on a skeleton and pull on the string to observe the movement produced.

Three laboratory sections were assigned to each of the cat dissection treatments; the clay sculpting treatment included four laboratory sections. In each treatment, three to four students worked together on a single anatomy representation. Students in the human clay sculpting treatment had no access to preserved cat specimens, and students in the two cat dissection treatments (no handout and handout) had no access to the human clay sculptures. All students had access to high-quality muscular and skeletal models (e.g., SOMSO, Coburg, Germany), anatomy atlases, and textbooks in their laboratory sections and were encouraged by their laboratory instructors to use these as part of their studies. Laboratory instructors were assigned to treatment groups so that there was a mix of graduate and undergraduate instructors with varying amounts of teaching experience. Students were asked to work within their treatment group for 1 wk before making requests to be reassigned to a different treatment group. After 1 wk, only one student made such a request. That student’s scores were eliminated from the analysis. There were no other student concerns regarding being part of an educational research study.

**Evaluation of Student Performance**

**Control exam scores.** Three lecture exams, which were a mix of higher- and lower-order questions, were administered during the semester. There were also two laboratory exams covering the skeletal system and the nervous/urogenital systems administered before and after the experimental portion of the course. These exams served as controls. For these units, all students had the same classroom experiences and answered identical questions. The average of the three lecture exam scores and the average of the two laboratory exam scores were calculated for each treatment group. Analyzing a series of control exam scores administered at different times ensures that the students assigned to each treatment group were of similar ability over the entire semester.

**Experimental exam scores.** The goal of this anatomy course, like many others, is to not only teach students the names and functions of the structures covered in class but also to equip students to use what they have learned in new settings that they may encounter in the future. With this in mind, this analysis focused on how well students were able to answer seven different types of higher-order anatomy questions (Table 1). The higher-order anatomy questions were not designed to reflect specific future scenarios for any one type of student but rather represented examples of scenarios that were not specifically covered during any of the classroom exercises and would therefore require students to apply the information they acquired in the classroom to a new situation. The higher-order anatomy questions were identical for all treatment groups. These questions required students to identify muscles presented from a novel point of view or to analyze functional anatomy problems and then describe the action of a muscle. Therefore, these questions were of greater difficulty than simply asking students to identify a pinned structure. Based on the results of our previous study (26), we expected the greater level of difficulty to increase our ability to discriminate treatment effects. These questions were either text or diagram based and administered during one of the 50-min lecture sections. All exam questions were reviewed by the course instructors. Students were told that they would be asked higher-order anatomy questions but were not given any additional instructions.

<table>
<thead>
<tr>
<th>Representation</th>
<th>Description</th>
<th>Expected Transformation Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human cadaver picture questions</td>
<td>Photograph of the posterior shoulder and back with arrows pointing to specific muscles</td>
<td>Cat dissection treatments: high; human sculpting treatment: low</td>
</tr>
<tr>
<td>Human cadaver cross-section picture questions</td>
<td>Photograph of a transverse section through the thigh with arrows indicating specific muscles</td>
<td>Cat dissection treatments: high; human sculpting treatment: low</td>
</tr>
<tr>
<td>Analysis of human limb movement questions</td>
<td>Diagram and text illustrating abduction and adduction of an upper limb</td>
<td>Cat dissection treatments: high; human sculpting treatment: low</td>
</tr>
<tr>
<td>Cat picture questions</td>
<td>Diagram of an undissected cat asking students to identify muscles with only external anatomy cues</td>
<td>Cat dissection treatments: high; human sculpting treatment: low</td>
</tr>
<tr>
<td>Analysis of cat limb movement questions</td>
<td>Diagram and text illustrating flexion and extension of an anterior limb</td>
<td>Cat dissection treatments: high; human sculpting treatment: low</td>
</tr>
<tr>
<td>Dissected horse diagram questions</td>
<td>Diagram showing a lateral view of superficial horse muscles</td>
<td>Cat dissection treatments: high; human sculpting treatment: low</td>
</tr>
<tr>
<td>Dissected frog diagram questions</td>
<td>Diagram showing a ventral view of superficial frog muscles</td>
<td>Cat dissection treatments: high; human sculpting treatment: low</td>
</tr>
</tbody>
</table>
Evaluation of Student Attitudes

Immediately before the muscular system unit, and before they knew which approach would be used in their laboratory section, students were asked a pair of questions to quantify their attitudes about the following three facets of the anatomy laboratory:

1. The general usefulness of performing cat dissections versus studying anatomy models in a teaching laboratory;
2. The specific importance of a dissection experience when learning both the name and function of anatomic structures; and
3. Their general attitude toward anatomy courses.

Likert scales were used for items 1 and 2. Lower numbers on the Likert scale represented a preference for dissection and higher numbers a preference for using models. The questions for item 3 were simple yes/no questions and were analyzed on a binomial scale, where a value of 1 was assigned to “yes” and 0 to “no”. Within each of the three items listed above, the students’ responses for each pair of questions were averaged. Students were asked the same questions immediately after the muscular system unit and at the end of the course after everyone had a chance to work with both models and dissected specimens.

Data Analysis

Reliability. A statistical reliability coefficient (α) was calculated for the scores from each exam and, when appropriate, from scores generated by subsets of exam questions. A reliability coefficient is an estimate of the stability, or reproducibility, of a set of scores and can range from 0.0 to 1.0. A reliability coefficient of 0.5 or greater is generally considered adequate (18). When calculating the α reliability coefficient for a subset of exam scores, it is likely that the calculated reliability for the subset will be lower than that of the overall exam. This is because the subset of questions has fewer questions than the overall exam. If the α reliability coefficient of a subset of questions is low but the questions themselves are not problematic, then the Spearman-Brown prophecy formula can be used to estimate what the α reliability coefficient would have been if there had been a greater number of questions (23).

Power analysis. It is important for any experiment to have adequate statistical power to detect a treatment effect. Experiments with low statistical power are more likely to mistakenly accept a false null hypothesis (detect no treatment effect, when in fact there was a treatment effect). By convention, a statistical power of at least 0.8 is considered necessary to confidently accept the null hypothesis that the difference between treatment groups is statistically insignificant (15). The statistical power of an experiment is determined by the number of subjects, the probability of rejecting a true null hypothesis (α), and treatment effect size. In this experiment, we had 222 students and set the probability of rejecting a true null hypothesis (α) to 0.05. We set the treatment effect size (Cohen’s d) to 0.5. Cohen considered this indicative of a medium-sized treatment effect, which he described as a difference visible to someone that is familiar with the parameter (3). Very large sample sizes can make it possible to detect very small treatment effects, but such differences may not be of practical significance to a classroom instructor. For this experiment, we defined a medium treatment effect to be of practical significance. Using the parameters defined above, we estimated the statistical power of our study to be 0.84 (5).

Control anatomy exams. Single-factor ANOVA was performed on the lecture and laboratory control anatomy exam scores for each of the three treatment groups to test for a significant treatment effect. If any effects were identified, pair-wise comparisons using t-tests were conducted as a followup.

Experimental anatomy exams. For the scores on the muscular system exam, student performance was evaluated for each of the seven subsets of question (Table 1) by calculating the mean exam score for each of the three treatment groups and then performing three pair-wise comparisons: 1) between the traditional cat dissection (no supplemental handout) and the human clay sculpting treatment; 2) between the supplemented cat dissection (who used a supplemental handout) and the human clay sculpting treatment; and 3) between the two cat dissection treatments (without and with a supplemental handout). Given that the previous study by Waters et al. (26) identified significant treatment effects, we chose to calculate a t-test statistic for all comparisons (using an unbiased estimate of SD). To test the null hypothesis that the difference between each pair of average scores equals zero, a P value was calculated, and the Bonferroni equation used to calculate the significance level for each comparison.

Student attitude surveys. For the student attitude surveys, the average response for each question was calculated within each treatment group, and the null hypothesis that the difference between the treatment means equals zero was then tested as described above.

RESULTS

Control Exam Scores

To ensure that there was no bias in the assignment of students to the three treatment groups, nor to the individual sections, the averaged scores of three lecture exams (α reliability coefficients: 0.7–0.8) and two laboratory exams (α reliability coefficients: 0.9 each) that were not included in the experimental portion of the course were used as controls (Fig. 1). When the averaged scores of the three lecture exams for each treatment group were compared with one another using single-factor ANOVA, no significant difference was detected among the treatment groups. Similarly, when the averaged scores of the two laboratory exams for each treatment group were compared with one another using single-factor ANOVA, no significant differences were detected among the treatment
Experimental Exam Scores

The muscular system exam comprised higher-order questions consisting of novel concepts and relationships that the students had not previously encountered in the classroom. The exam was divided into seven subsets of questions (Table 1). Even though the overall α reliability coefficient for the entire exam was strong (0.9), α reliability coefficients were also calculated for each of the seven subsections and, with one exception, ranged from 0.6 to 0.9. The α reliability coefficient for the questions asking students to identify muscles on a horse anatomy diagram was extremely low (0.3). Since there were a limited number of questions in each subset, the Spearman-Brown prophecy formula was used to estimate what the α reliability coefficient would have been if there were 10 questions in each subset. After a Spearman-Brown adjustment for 10 questions/substrate, the estimated reliability coefficients for all 7 question subsets ranged from 0.6 to 0.9.

Human higher-order anatomy questions. Figure 2 shows the scores for the three types of human higher-order anatomy questions. When asked to identify muscles on a photograph of a dissected human cadaver, students in the human clay sculpting treatment earned significantly higher scores (mean: 61% and SD: 34%, n = 84) than the students who performed cat dissections with no handout (mean: 31% and SD: 25%, n = 69) and with the handout (mean: 32% and SD: 27%, n = 69). There were no significant differences between the cat dissection treatments with and without the handout. When asked a similar set of identification questions using a diagram of a transverse section through the middle of the thigh, students in the human clay sculpting treatment again earned significantly higher scores (mean: 54% and SD: 41%, n = 84) than the students who performed cat dissections with no handout (mean: 37% and SD: 36%, n = 69) or with a handout (mean: 37% and SD: 38%, n = 69). However, when students were asked to analyze functional anatomy questions about human muscle actions there were no significant differences in test scores between the cat dissection with no handout group (mean: 24% and SD: 22%, n = 69), the cat dissection with a handout group (mean: 24% and SD: 24%, n = 69), or the human clay sculpting group (mean: 28% and SD: 22%, n = 84). In two of three measures, the human clay sculpting experience had a significantly positive effect on students’ ability to answer higher-order human anatomy questions.

Cat higher-order anatomy questions. Figure 3 shows the scores for two types of higher-order cat anatomy questions. When asked to identify muscles from an external anatomy diagram of a cat, students in the human clay sculpting treatment earned significantly poorer scores (mean: 18% and SD: 35%, n = 84) than students who performed cat dissections with no handout (mean: 52% and SD: 35%, n = 69) or with a handout (mean: 50% and SD: 36%, n = 69). When students were asked to analyze functional anatomy questions about cat muscle actions, students assigned to the human clay sculpting treatment again earned poorer scores (mean: 32% and SD: 28%, n = 84) than students who performed cat dissections with no handout (mean: 44% and SD: 33%, n = 69) or with a handout (mean: 38% and SD: 32%, n = 69), but only the comparison between the human clay sculpting group and the cat dissection with no handout was significant. The presence or absence of a handout during a cat dissection had no measurable effect on student performance. Participating in a cat dissection clearly helped students identify muscles on a cat diagram and seemed to have some positive effect when analyzing functional anatomy questions about cat muscle actions.

Horse and frog higher-order anatomy questions. In addition to the human and cat anatomy questions, students were asked to identify muscles on organisms that neither group had studied in the laboratory (Fig. 4). On questions requiring students to identify muscles on a horse cadaver diagram, students in the human clay sculpting treatment performed significantly worse (mean: 10% and SD: 13%, n = 84) than both the students who performed cat dissections with no handout (mean: 24% and SD: 22%, n = 69) or with a handout (mean: 25% and SD: 22%, n = 69). Once again, there were no significant differences between the two cat dissection treatments.

1 Supplemental Material for this article is available at the Advances in Physiology Education website.
Students had a chance to work with both models and dissected cats. There were no differences detected in student attitudes between the two cat treatments for any question. When students were asked about their general attitude toward anatomy (Table 4), there were no treatment effects detected. Students show clear preferences toward the specific approach used in their anatomy class when asked to rate one approach versus the other, but a cat dissection versus human clay sculpting experience did not seem to affect how they felt about anatomy as a field of study.

DISCUSSION

When designing a human anatomy course, anatomy educators may choose to teach from a variety of anatomic representations including models, computer programs, human cadavers, and animal cadavers. On questions requiring students to identify muscles on a frog cadaver diagram, there were no significant differences in test scores between the human clay sculpting group and the cat dissection groups without or with a handout. This response pattern continued even to the conclusion of the course, after all students had a chance to work with both models and dissected cats. There were no differences detected in student attitudes between the two cat treatments for any question. When students were asked about their general attitude toward anatomy (Table 4), there were no treatment effects detected. Students show clear preferences toward the specific approach used in their anatomy class when asked to rate one approach versus the other, but a cat dissection versus human clay sculpting experience did not seem to affect how they felt about anatomy as a field of study.

**Fig. 3.** Laboratory exam scores (means ± SE) for two types of higher-order cat anatomy questions. **Left:** when asked to identify structures on a cat external anatomy diagram, students in the human clay sculpting treatment performed significantly worse than those in both the cat dissection treatment without a handout (P < 0.000001) and with a handout (P < 0.000001). There were no differences between the two cat dissection treatments (P = 0.68). **Right:** when asked to analyze a functional anatomy question on muscle action of a cat, students in the cat dissection treatment that did not use a handout performed significantly better than their classmates assigned to the human clay sculpting treatment (P = 0.02), but no other significant treatment effects were detected between either the cat dissection treatments (P = 0.34) or cat dissection using a handout versus the human clay sculpting treatment (P = 0.21).

On questions requiring students to identify muscles on a horse cadaver diagram, there were no significant differences in test scores between the human clay sculpting group (mean: 52% and SD: 34%, n = 84), the cat dissection without a handout group (mean: 46% and SD: 33%, n = 69), or the cat dissection with a handout group (mean: 46% and SD: 32%, n = 69). Performing a cat dissection helped students answer questions on horse anatomy, but all three treatment groups performed similarly when attempting to answer questions on frog anatomy.

**Student Attitudes**

Immediately before the muscular system unit, immediately after the muscular system unit, and then at the conclusion of the course, students were asked three types of questions relating to their anatomy experiences: to indicate their general preference for studying anatomy by performing a cat dissection versus using anatomic models in the classroom (Table 2), to rate the importance of a cat dissection experience versus anatomy models when the course goal was to learn both the names and functions of anatomic structures (Table 3), and, finally, to share their general attitude toward anatomy as a field of study (Table 4). When asked their opinions before the experimental portion of the course, there were no differences detected between any of the treatment groups to any of the questions, indicating that the assignment to each of the treatment groups was unbiased with regard to student attitude. By the midpoint of the course, when asked about their preferences for performing dissections versus using models (Table 2) and the value of dissection versus studying models when learning anatomic names and functions (Table 3), both groups performing cat dissections preferred the dissection approach significantly more than the group building human clay sculptures. This response pattern continued even to the conclusion of the course, after all students had a chance to work with both models and dissected cats. There were no differences detected in student attitudes between the two cat treatments for any question. When students were asked about their general attitude toward anatomy (Table 4), there were no treatment effects detected. Students show clear preferences toward the specific approach used in their anatomy class when asked to rate one approach versus the other, but a cat dissection versus human clay sculpting experience did not seem to affect how they felt about anatomy as a field of study.

**Fig. 4.** Laboratory exam scores (means ± SE) for two types of higher-order nonhuman/noncat anatomy questions. **Left:** when asked to identify structures on a horse cadaver diagram, students in the human clay sculpting treatment performed significantly worse than those in both the cat dissection treatment without a handout (P = 0.000003) and with a handout (P = 0.000001). There were no differences between the two cat dissection treatments (P = 0.92). **Right:** when asked to analyze a functional anatomy question on muscle action of a cat, students in the cat dissection treatment that did not use a handout performed significantly better than their classmates assigned to the human clay sculpting treatment (P = 0.02), but no other significant treatment effects were detected between either the two cat dissection treatments (P = 0.34) or cat dissection using a handout versus the human clay sculpting treatment (P = 0.21).
and dissected cats. Waters et al. (26) reported that students who studied anatomy by sculpting a human representation from clay were more successful answering anatomy exam questions than students who dissected cats. This study tested whether Waters et al.’s (26) finding that a clay sculpting experience yielded better test scores than a cat dissection lesson could be attributed to properties of clay sculpting, or specifically to a classroom handout effect, or perhaps to the similarity between the human anatomy representations used in the classroom versus those on the anatomy exam.

Effect of Classroom Handouts

The first question tested in this study was the hypothesis that treatment differences reported by Waters et al. (26) on the posttest were caused by differences in the instructional materials provided to the two groups. Recall that in the Waters et al. (26) study, not only did experimental groups differ with respect to the assigned lab activities, but also that supplemental handouts were provided to the clay sculpting group but not the cat dissection group. We tested the viability of this explanation by including a third treatment to which students who completed cat dissections were provided a parallel supplemental handout. We predicted that if the effects of the Waters et al. (26) study were due to difference in the provided instructional materials, then students in the two groups with the handout would obtain higher posttest scores than participants without the handout. We did not find this effect. First, there were no detectable differences between the group of students who completed a cat dissection with the handout and those who completed cat dissection without the handout. Second, the pattern of results revealed significant differences between the groups of students who completed cat dissections with the handout and those who completed the human clay sculpting activity with the handout. This demonstrates that the student performance differences between a traditional cat dissection experience and a human clay sculpting experience reported by Waters et al. (26) are not attributable to the human clay sculpting treatment using a supplemental handout versus the cat dissection having no handout.

Similarity/Dissimilarity of Anatomy Representations

The second explanation tested concerns the match between the anatomic representations that students studied in the laboratory versus the representations used on the exam. Although students who engaged in the human clay sculpting activity in the Waters et al. (26) study obtained higher posttest scores than did those who completed cat dissections, these differences may have been related to the fact that the posttest used by Waters et al. (26) primarily contained questions that used human anatomic representations. In this study, we tested this explanation by varying the representations used on the posttest questions. Students were asked to identify structures from human images, cat images, and even horse and frog images.

Table 2. Attitudes toward dissection versus anatomy models in human anatomy teaching laboratories

<table>
<thead>
<tr>
<th>Survey</th>
<th>Cat Dissection Group With No Handout (Group 1)</th>
<th>Cat Dissection Group With a Handout (Group 2)</th>
<th>Human Clay Sculpting Group (Group 3)</th>
<th>P Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Means ± SE No. of students</td>
<td>Means ± SE No. of students</td>
<td>Means ± SE No. of students</td>
<td></td>
</tr>
<tr>
<td>Preexperiment</td>
<td>2.94 ± 0.14 62</td>
<td>2.93 ± 0.12 59</td>
<td>2.92 ± 0.10 74</td>
<td>0.92, group 1 vs. group 3; 0.93, group 2 vs. group 3; and 0.99, group 1 vs. group 2</td>
</tr>
<tr>
<td>Midexperiment</td>
<td>2.78 ± 0.16 47</td>
<td>2.51 ± 0.15 51</td>
<td>3.22 ± 0.13 58</td>
<td>0.04, group 1 vs. group 3; 0.0006, group 2 vs. group 3; and 0.23, group 1 vs. group 2</td>
</tr>
<tr>
<td>Postexperiment</td>
<td>2.38 ± 0.13 51</td>
<td>2.53 ± 0.16 49</td>
<td>3.14 ± 0.13 69</td>
<td>0.0001, group 1 vs. group 3; 0.003, group 2 vs. group 3; and 0.48, group 1 vs. group 2</td>
</tr>
</tbody>
</table>

Scores were evaluated on a scale of 1-5, where 1 = dissections are the best/most helpful and 5 = anatomy models are the best/most helpful. In the surveys, the scores of the following two Likert-scale questions were averaged: “If you were in charge of this course and could use dissection and/or anatomy models to teach your students, what would you use to teach your students anatomy?” and “As an anatomy student, is it more important to study anatomy with a dissection or is using anatomy models more important?” P values indicate differences between treatment groups within each survey. See METHODS for calculations.

Table 3. Attitudes toward the importance of dissection in human anatomy teaching laboratories

<table>
<thead>
<tr>
<th>Survey</th>
<th>Cat Dissection Group With No Handout (Group 1)</th>
<th>Cat Dissection Group With a Handout (Group 2)</th>
<th>Human Clay Sculpting Group (Group 3)</th>
<th>P Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Means ± SE No. of students</td>
<td>Means ± SE No. of students</td>
<td>Means ± SE No. of students</td>
<td></td>
</tr>
<tr>
<td>Preexperiment</td>
<td>3.15 ± 0.14 61</td>
<td>3.25 ± 0.11 59</td>
<td>3.20 ± 0.11 74</td>
<td>0.75, group 1 vs. group 3; 0.74, group 2 vs. group 3; and 0.55, group 1 vs. group 2</td>
</tr>
<tr>
<td>Midexperiment</td>
<td>2.65 ± 0.15 47</td>
<td>2.78 ± 0.15 51</td>
<td>3.32 ± 0.13 58</td>
<td>0.001, group 1 vs. group 3; 0.009, group 2 vs. group 3; and 0.53, group 1 vs. group 2</td>
</tr>
<tr>
<td>Postexperiment</td>
<td>2.70 ± 0.14 51</td>
<td>2.71 ± 0.15 49</td>
<td>3.18 ± 0.13 69</td>
<td>0.02, group 1 vs. group 3; 0.03, group 2 vs. group 3; and 0.93, group 1 vs. group 2</td>
</tr>
</tbody>
</table>

Scores were evaluated on a scale of 1-5, where 1 = dissections are the best/most helpful and 5 = anatomy models are the best/most helpful. In the surveys, the scores of the following two Likert-scale questions were averaged: “If you were in charge of this course and could use dissection and/or anatomy models to teach your students, what would you use to teach your students anatomy?” and “As an anatomy student, is it more important to study anatomy with a dissection or is using anatomy models more important?” P values indicate differences between treatment groups within each survey. See METHODS for calculations.
While studying anatomy, students must select the relevant parts of the visual representation to create a mental model of the material they are trying to learn (13). This mental model is more than simply recalling a visual image, because the model also integrates names, facts, and other concepts the student believes to be related to the image (12). While studying, the relevant content of the representation and the connections the student makes to other knowledge will both help the student create an accurate mental model and promote learning. In contrast, extraneous information or requiring the student to process the representation in ways that are beyond him/her will interfere with learning (19).

In our study, student performance was higher on questions that did not require students to transform an image too extensively. Students were more successful answering identification questions when the representation used in the exam question was similar to the representation studied in the laboratory. For example, students in the human clay sculpting treatment obtained higher scores when answering human cadaver questions than students who studied cat dissection. Conversely, participants who studied a cat dissection obtained the highest scores on cat anatomy questions. The same pattern was observed when students were asked to identify muscles of an adult horse. The horse muscle diagram presented muscles from a lateral view, and given that cats and horses are both mammalian quadrupeds, the shape and organization of the horse muscles were likely more similar to cat specimens than to human specimens, requiring less transformation of the mental model for the students performing cat dissections and more transformation for the students sculpting human muscles from clay. No differences in exam scores between any of the groups were detected when students were asked to identify muscles on a frog diagram. For students enrolled in an introductory anatomy course, amphibian anatomy may be perceived as so different from that of both cat and human that the amount of transformation for each treatment group was similar, resulting in no detectable treatment effects. Our data show that different representations of anatomic structures affect student performance. Similarly, Schnotz and Bannert (21) demonstrated that alternative representations of a concept can lead to differences in posttest performance, even when the representations are informationally equivalent. In their study (21), students used alternate representations of the earth’s time zones to answer different types of questions requiring time zone calculations. They reported that while alternate representations may contain identical information, they still may not be equally effective ways to learn specific concepts. The representations used in the cat dissection and human clay sculpting treatments can be considered informationally equivalent because the content available is the same in both (9). Our data indicate that students have a difficult time performing transformations from one representation of anatomy to another. One must note that our data does not indicate that one representation is necessarily better than the other. Rather, which representation is best is determined by the task for which it is needed. If one wishes to learn human anatomy, then studying a human anatomic representation is best (14). If one wishes to learn cat anatomy, then studying a cat representation is better than an alternative representation.

### Human and Cat Functional Anatomy

When students were presented with a novel functional anatomy question and then asked to identify a muscle’s action, there was a nonsignificant trend favoring students assigned to the human clay sculpting treatment when the questions were presented in a human anatomy context and a similar trend favoring the students assigned to the cat dissection treatments when the questions were presented in a cat anatomy context. This trend is consistent with our hypothesis that student performance will improve when the studied representation is similar to the form of the tested representation, since similar representations will require fewer mental transformations. Only one comparison, however, detected a significant treatment effect: when answering a cat muscle action question, the cat dissection treatment with no supplemental handout performed better than the human clay sculpting treatment. The muscle action questions used in this study were novel to the students and thus could not be answered by rote. To answer such a question, a student must transform the information presented by a single image on the exam into a mental animation of muscle action. The transformations that need to occur to generate this mental animation may be difficult enough to prevent consistently measuring a treatment effect under the conditions of this experiment. Hegarty et al. (5) demonstrated that mentally animating a process from a single static image is the most difficult type of transformation. More data are necessary to make any conclusive statement about the effect of a cat dissection versus human clay sculpting experience on students’ ability to answer muscle action questions.
Student Attitudes

When students were asked about their general preferences regarding studying from models versus performing cat dissections (Table 2) and for their opinions about the value of performing dissections when they need to learn both the name and function of a structure (Table 3), there were no significant treatment effects before the experimental portion of the course, demonstrating again that there was no bias in the assignment of the individual sections to the three treatment groups. However, when students were asked the same questions immediately after completing the muscular system unit and then a third time at the end of the semester (after everyone had a chance to work with both models and dissected cats), students demonstrated a significant preference toward the type of specimen used in their treatment group. This result is consistent with most other studies that have asked similar questions (2, 17, 22, 26) and suggests that student attitudes toward dissection versus models alone should not determine the specific approach an instructor chooses for the classroom, since, on average, students will likely come to value whichever approach they experience.

When students were asked about their general attitudes toward anatomy courses (Table 4), there were no detectable treatment effects at any point during the semester. This result is consistent with that reported by Waters et al. (26) when a human clay sculpting lesson was compared with a cat dissection experience but differs from a separate unpublished comparison between a human clay sculpting experience and a prospected human cadaver lesson (see ACKNOWLEDGEMENTS). In the latter experiment, students that studied a human cadaver in class were significantly more positive about anatomy courses than their classmates who sculpted human structures from clay. The affective experience of working with a human cadaver may be distinct from working with models or even dissecting other organisms. Additional research is required to demonstrate whether or not such an experience may improve student learning.

Conclusions

In an earlier study (26), we showed that students sculpting human structures from clay earned higher anatomy exam scores compared with students performing cat dissections. In the present work, the reason for this difference was investigated. The data discount the influence of handouts along with the assignment of the individual sections to the three treatment groups (Table 2) and for their opinions about the value of dissection versus models tend to be fluid, and students will generally come to value whichever approach the instructor chooses.

ACKNOWLEDGMENTS

The authors thank the faculty and administration at The Pennsylvania State University (Penn State) and Mohawk Valley Community College (MVCC) for encouragement and support, especially Dr. Robert Jubenville at MVCC and the Institutional Review Board (IRB) at Penn State who reviewed and approved this study (IRB no. 28174). The authors also thank all of Penn State’s anatomy students and teaching assistants for cooperation and hard work and Dominic Minichiello for technical assistance.

Unpublished data cited in this article is from a study presented at the 2006 Annual Meeting of the Human Anatomy and Physiology Society (J. Waters, S. Drogo, W. Perrotti, and D. Kelly. Prosected cadavers vs. clay sculpting: a comparison of two approaches to an undergraduate human anatomy laboratory.)

Finally, the authors dedicate this article to the memory of our dear friend and colleague, Sam Drogo (1948–2010).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

REFERENCES