Improved Understanding of Human Anatomy through Self-guided Radiological Anatomy Modules

Andrew W. Phillips, MEd, Sandy G. Smith, PhD, Callum F. Ross, PhD, Christopher M. Straus, MD

Rationale and Objective: To quantifiably measure the impact of self-instructed radiological anatomy modules on anatomy comprehension, demonstrated by radiology, gross, and written exams.

Materials and Methods: Study guides for independent use that emphasized structural relationships were created for use with two online radiology atlases. A guide was created for each module of the first year medical anatomy course and incorporated as an optional course component. A total of 93 of 96 eligible students participated. All exams were normalized to control for variances in exam difficulty and body region tested. An independent t-test was used to compare overall exam scores with respect to guide completion or incompletion. To account for aptitude differences between students, a paired t-test of each student’s exam scores with and without completion of the associated guide was performed, thus allowing students to serve as their own controls.

Results: Twenty-one students completed no study guides; 22 completed all six guides; and 50 students completed between one and five guides. Aggregate comparisons of all students’ exam scores showed significantly improved mean performance when guides were used (radiology, 57.8% [percentile] vs. 45.1%, \( P < .001 \); gross, 56.9% vs. 46.5%, \( P = .001 \); written, 57.8% vs. 50.2%, \( P = .011 \)). Paired comparisons among students who completed between one and five guides demonstrated significantly higher mean practical exam scores when guides were used (radiology, 49.3% [percentile] vs. 36.0%, \( P = .001 \); gross, 51.5% vs. 40.4%, \( P = .005 \)), but not higher written scores.

Conclusions: Radiological anatomy study guides significantly improved anatomy comprehension on radiology, gross, and written exams.

Key Words: Undergraduate medical education; anatomy; radiology; self-guided instruction; radiological anatomy; basic science education.

©AUR, 2012

Concern over the inadequacy of preclinical medical education for residency preparation has led to reevaluation of teaching methods in many content areas. Undergraduate anatomy education has been specifically identified as problematic because students and residents currently have difficulty applying their anatomical knowledge to clinical situations (1–4). These concerns, along with new, more readily available technologies, have led to a movement to use radiologic imaging technologies in anatomy instruction. For example, reports of ultrasound images of visceral organs in live models, computerized image databases, plain film images in the cadaver rooms, and small- and large-group lectures that use digital radiographic images projected on monitors are typical of the technological advances disseminated in preclinical medical anatomy classes (5–7).

“Blended learning” refers to the incorporation of various teaching modalities for the same endpoint—such as modern medical imaging combined with traditional gross dissection to comprehend anatomy. Previous research has found support for the use of blended learning in anatomy. For instance, when used as an adjunct to cadaver dissection, Shaffer reported that two-dimensional imaging resulted in positive student feedback, and Pereira found that annotated nonradiologic graphics resulted in improved performance on anatomy exams (8,9). Multiple representations of anatomical structures thus appear to help students understand anatomy.

Clinical relevance of material has also been associated with higher quality learning, which suggests that incorporating clinically relevant medical imaging with gross cadaveric anatomy may provide an enhanced learning experience for students (10). However, research has been generally qualitative in nature, such as reporting higher student satisfaction in courses that integrate radiology and anatomy (10,11). Most quantitative research has focused on the impact of an integrated course of radiology and anatomy on radiographic image interpretation that may not be a complete measure of anatomical comprehension (6,12–14). Recently, quantitative improvement using unlabeled serial imaging was observed only in head and neck practical performance (combined gross and radiological questions) and in a subset of
radiological questions designed to test spatial reasoning for all body sections combined (15). Thus, despite the growing allocation of financial and personnel resources to the integration of radiology with anatomy, the extent to which radiographic imaging improves conceptual comprehension of anatomy is not well established empirically.

In this study, we investigate the utilization of serial and still imaging with a corresponding instruction guide during a preclinical gross anatomy course. On the basis of previous research, we predict that the use of blended, clinically relevant imaging during a gross anatomy course will be positively linked to enhanced anatomical structure knowledge, as demonstrated by radiology, gross, and written exams.

MATERIALS AND METHODS

Study Setting and Participants

All protocols for this study were granted exemption status by the University of Chicago institutional review board, and written informed consent was obtained from all participants as an additional measure of ethical consideration. A prospective, quasi-experimental study was performed with participants recruited from the 102-member 2008 first-year class of an urban university teaching hospital. The course spanned two academic quarters. All participation was voluntary, blinded to instructors, and had no relationship with course grades. Five Medical Scientist Training Program students and one PhD student who began the course early without the study guide intervention were excluded from analysis.

Study Design

Self-administered study guides were created in Word 2007 (Microsoft Corporation, Redmond, WA) for use with two online radiology atlases, E-Anatomy and NetAnatomy (16,17). Neither publisher was consulted, nor was funding received from either for this investigation. Study guide components included specific structures to locate, descriptions of their locations and relationships to other structures, and questions (with answers) applying spatial reasoning and anatomic awareness (Table 1). Guides were almost entirely text with one to two self quiz images at the end. Each guide of approximately two hours duration (except for an expanded head and neck 6 hour duration study guide) was created by radiology faculty and second-year medical student teaching assistants for each of the six body regions (thorax, abdomen, pelvis, head and neck, upper limb, and lower limb) and was offered in parallel with the first-year medical student anatomy course. Image modalities included computed tomography (CT), CT angiogram, magnetic resonance (MR), x-ray, angiogram, ultrasound, and echocardiogram, with emphasis on serial planar CT and MR imaging. The study guides were an optional, out-of-classroom component of the 2008 human morphology course. Students self-reported which guides they completed in a survey at the conclusion of the course. The required course components for each module included a 1-hour radiology lecture, daily 1-hour anatomy lectures, daily 20-minute radiologic and cadaveric structure correlation small group sessions, a 30-minute radiological anatomy small group session, a gross anatomy and radiological anatomy review, and daily gross anatomy laboratory. Imaging modalities and structures visualized were similar between the study guides and required course components. Self-instruction was the critical pedagogical difference between the study guides and other course components.

Radiology practical exams composed of 18–24 short answer questions each were written by radiology faculty and edited jointly with anatomy faculty. All exam images were obtained from sources other than the atlases used for the guides and were of the aforementioned image modalities. Gross practical (50 short-answer questions each) and written exams (text and diagram short answer, matching, and multiple choice questions) were created exclusively by anatomy faculty who were blinded to final versions of the study guides. Radiology and gross practical exam questions tested multiple depths of anatomical spatial relationship understanding, from identifying visible structures (eg, a bone on an x-ray) to inferring nonvisualized structures (eg, insertion location of a muscle on a skeleton) to sequencing serial images. A revised version of Bloom’s Taxonomy of Learning Objectives (18) was used as a template to create radiology exam questions that tested different depths of learning based on increasingly demanding comprehension of structures’ spatial and physiological relationships. For example, some questions included structures that were not themselves clearly visualized in the image but could be inferred from the clearly visible related structures if students understood spatially what should be in a given location in a given plane. Exam questions were quantitatively validated in a separate analysis that demonstrated five distinct categories of questions suggestive of hierarchical levels with moderate reliability (19). In contrast to the spatial relationship emphasis of the practical exams, written exam questions more commonly emphasized physiology (eg, movement actions and innervations of limb muscles). Scores from all three exam types were added into a single grade for each module for the purposes of student course grades (not reported here).

Data Analysis

To control for variation in exam difficulty, student scores were normalized (20) around the class mean for each of the 18 total exams (three exam types for each of six body regions). This allowed for a normalized score for each student on each exam type for each body region. These normalized scores were used in all subsequent analyses. With variation in exam difficulty thus controlled, each student’s performance could then be characterized by his or her use (or not) of a study guide. A two-tailed independent group t-test was performed on all normalized scores associated with use or nonuse of
TABLE 1. Characteristic Examples of Study Guide Instructions

<table>
<thead>
<tr>
<th>From Abdomen, X-ray Imaging</th>
<th>From Thorax, Serial Computer Tomography</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Image #10: Barium contrast, anteroposterior view</strong></td>
<td><strong>Click on the “Anatomical Structures” button and click on “Bones.”</strong></td>
</tr>
<tr>
<td>Note the characteristics of the large intestine:</td>
<td>View the ascending aorta until it reaches T4/T5. Look to the left of the</td>
</tr>
<tr>
<td>Omental appendices, taenia coli, haustra.</td>
<td>image and appreciate the level where the slice was taken with respect</td>
</tr>
<tr>
<td>Which parts of the colon are intraperitoneal and</td>
<td>to lung and heart positions. In a single section, you should be able to see</td>
</tr>
<tr>
<td>secondarily retroperitoneal?</td>
<td>the following characteristic landmarks:</td>
</tr>
<tr>
<td>Intraperitoneal: cecum, appendix, transverse colon,</td>
<td>-beginning and end of the aortic arch</td>
</tr>
<tr>
<td>sigmoid colon</td>
<td>-bifurcation of the trachea</td>
</tr>
<tr>
<td>Secondarily retroperitoneal: ascending colon,</td>
<td>-pulmonary trunk at its most superior point</td>
</tr>
<tr>
<td>descending colon</td>
<td></td>
</tr>
</tbody>
</table>

a study guide for each exam type (radiology practical, gross practical, and written) to determine if there was any overall difference in scores as a result of using study guides.

To control for varying student aptitudes and study habits contributing to between-subject variance, exam scores were also analyzed with a two-tailed paired *t*-test with respect to each student’s performance with and without use of the study guide, thus allowing students to serve as their own controls. To perform this analysis, each student’s normalized exam scores were averaged across exams for which a study guide was used and across exams for which a study guide was not used. For this paired analysis, the data from students who completed all six study guides or no study guides had to be excluded as they were only able to provide a score for one condition.

For both analyses, Pearson’s correlation coefficient was calculated to determine intervention effect sizes (21).

All data was entered into Excel 2007 (Microsoft Corporation) and analyzed with SPSS version 16 (Statistical Package for the Social Sciences Corporation, Chicago, IL) with the exception of effect sizes, which were calculated and verified twice with TI-83 Plus (Texas Instruments, Dallas, TX).

RESULTS

Descriptive Data

Of the 96 students in the 2008 first-year medical student class who met inclusion criteria, 93 students responded, yielding a 96.9% response rate. Table 2 describes the respondents’ demographic data to provide context for our findings. A demographic comparison of those who did and did not complete study guides was implausible because students completed varied numbers of guides corresponding to various body region modules. Figure 1 displays the frequency distribution of completed guides by course module. Stratifying use by the number of guides that each student completed demonstrated a minimally bimodal pattern regardless of how many total guides students completed. Guide use was, in general, evenly distributed throughout the course by students who used varying numbers of the guides except for a bimodal increase at the start and middle of the course by students who completed relatively fewer study guides than their classmates.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years ± SD</td>
<td>23.5 ± 2.9</td>
</tr>
<tr>
<td>Male sex, number/total (%)</td>
<td>41/88 (46.6)</td>
</tr>
<tr>
<td>Bachelor’s degree completed within 3 months of beginning medical degree, number/total (%)</td>
<td>42/89 (47.2)</td>
</tr>
<tr>
<td>Bachelor’s degree in a basic science, number/total (%)</td>
<td>58/89 (65.2)</td>
</tr>
<tr>
<td>Previously attained advanced science degree, number/total (%)</td>
<td>9/89 (10.1)</td>
</tr>
<tr>
<td>Parent(s) trained as physician, number/total (%)</td>
<td>17/88 (19.3)</td>
</tr>
<tr>
<td>Ethnicity, number/total (%)</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>10/85 (11.8)</td>
</tr>
<tr>
<td>Asian American</td>
<td>23/85 (27.1)</td>
</tr>
<tr>
<td>Caucasian</td>
<td>43/88 (50.6)</td>
</tr>
<tr>
<td>Native American</td>
<td>3/85 (3.5)</td>
</tr>
<tr>
<td>Other</td>
<td>6/85 (7.1)</td>
</tr>
</tbody>
</table>

SD, standard deviation.

*Varying total n values are due to incomplete demographic information provided.

Aggregate Analysis

Twenty-one students completed no study guides, 22 completed all six guides, and 50 students completed between one and five guides (Table 3).

In comparing the aggregate scores obtained for exams with respect to use or nonuse of a study guide, scores were significantly higher on all three exam types when study guides were used (Table 3). The radiology practical exam means with and without study guide use were represented by the 57.8 percentile and 45.1 percentile, respectively (Z = 0.197 and Z = −0.124, respectively; *t*(545) = −3.87, *P* < .001). For the gross practical exam, average student performance was at the 56.9 percentile when the study guides were used and at the 46.5 percentile when they were not used (Z = 0.174 and Z = −0.0871, respectively; *t*(545) = −3.21, *P* = .001). Student means for the written exams were at the 57.8 percentile and the 50.2 percentile when study guides were and were not used, respectively (Z = 0.197 and Z = 0.004, respectively; *t*(544) = −2.54, *P* = .011). Scores throughout the class were
The intervention yielded an effect size of $t(49) = 2.95$, $(Z = 0.037, 51.5\ percentile\ vs.\ Z = P(0.005))$ when study guides were completed. The intervention yielded an effect size of $r = 0.44$ for the radiology practical exams and $r = 0.39$ for the gross practical exams. There was no significant difference in written exam scores with respect to whether or not a study guide was used (Fig 2).

Paired Analysis

Fifty students completed between one and five study guides and were able to serve as their own controls in a paired analysis of their scores with and without use of a study guide. These scores were compared using a paired scores $t$-test. Performance was significantly better on radiology exams ($Z = -0.019, 49.3\ percentile\ vs.\ Z = -0.359, 36.0\ percentile; t(49) = 3.39, P = .001$) and on gross practical exams ($Z = 0.037, 51.5\ percentile\ vs.\ Z = -0.242, 40.4\ percentile; t(49) = 2.95, P = .005$) when study guides were completed. The intervention yielded an effect size of $r = 0.44$ for the radiology practical exams and $r = 0.39$ for the gross practical exams. There was no significant difference in written exam scores with respect to whether or not a study guide was used (Fig 2).

**DISCUSSION**

In general, students performed significantly better on exams when they used the self-guided imaging instruction modules. Paired analyses allowing students to serve as their own controls reduced important confounders such as student aptitude and demographic effects. It also demonstrated statistically and practically significant improvement on both radiology and gross practical exams for which students used the study guides. The paired $t$-test effect sizes suggest the study guides had a medium effect on the gross practical exams (accounting for approximately 15% of the total variance) and a large effect on the radiology practical exams (accounting for approximately 20% of the total variance) (22). Of additional practical significance, those students who completed some but not all study guides performed below the class mean when the guides were not used and rose to the class mean when the guides were used.

We propose that improved spatial anatomical understanding played a central role in the observed exam score improvement. First, study guides intentionally emphasized spatial relationships. Additionally, score improvements were observed on exams that had been previously validated for deeper spatial relationship understanding (23). Unfortunately, because of the small class size our study was not powered to test differences in exam question taxonomy level by use or nonuse of the study guides. Nonetheless, the emphasis on spatial relationships in the instruction and assessment components creates a cohesive explanation for the improved anatomy comprehension, especially across cadaveric and radiologic structures.

Our results are also consistent with previous studies that demonstrated improved learning when teaching methods incorporated clinically relevant material and a blend of approaches (8–10). Conversely, our present findings differ from a similar study that tested instructor-taught sessions that described spatial relationships between structures on cadavers and images in which no difference on exams was observed (23). This may suggest a particular benefit to self-instructed learning in anatomy.

The present study additionally provides empirical support for the long-held concept that radiological instruction can enhance anatomical comprehension (24). It is well-established that reasoning and visualization are superior to memorization in understanding and retaining anatomy (25,26). Given that reasoning and visualization are critical for interpreting radiographic images, it is plausible that radiology could provide an additional, unique method of conveying anatomical understanding. Moreover, the improvement on the gross practical exams as a result of studying imaging suggests that, for at least some students, imaging can fulfill a learning need that is not otherwise met in traditional anatomy instruction.

The improvement in the radiology and gross practical scores as a result of an imaging intervention suggests that rote memorization of images did not account for the gross practical findings. Moreover, the gross exams were created by faculty who were blinded to the final versions of the study guides, so gross practical exam questions were not paired with questions in the study guides. Furthermore, none of the images referenced in the study guides was used for the radiology exams, suggesting that the radiology exam improvement is more indicative of successful image interpretation than rote image recall. Because conceptualizing three-dimensional structures in two-dimensional images can be particularly challenging for many students, using the study guides concurrently with cadaver dissection may have helped students understand the fundamentals of image interpretation and...
thus improve their radiology exam scores (24). The ability to conceptually translate structural relationships in both gross and imaging contexts lends further support to structural spatial relationship understanding as an important source of exam improvement. Moreover, it suggests deeper learning and would likely have positive implications for students’ preparedness for clinical rotations.

Interpretation of below-mean exam scores rising to the class mean when students used some of the study guides should be made with caution. This improvement trend suggests the guide may be an invaluable tool for educators to reach a population of students that may be struggling with anatomy as traditionally taught. It has been posited that defects in visual memory contribute to poor performance in anatomy (27). If that assertion is accurate, it is plausible that the study guides’ emphasis on spatial understanding rather than visual memorization accounts for the distribution of exam performance improvement.

Although the aggregate analysis demonstrated improvement on all three exam types when students used the study guides, the paired analysis showed improvement only on the radiological and gross practical exams. Important personal trait confounders were likely present in the aggregate analysis, such as individual student aptitude, motivation, and study habits. The possibility that improved aggregate exam scores with study guide use may simply be a marker for “better” students who studied more was, interestingly, not supported by the exam scores stratified by number of study guides used. There was not a clear trend of increased scores as the total number of study guides increased. Confounders were nonetheless present in the aggregate analysis, and the paired analysis mitigated the effects of these largely immeasurable between-student confounders because students served as their own controls. Therefore the positive findings in the radiographic and gross practical exams provide more compelling evidence than the aggregate data for the effects of radiographic instruction. Additionally, no difference may have been observed on the written exams in the paired analysis because the written exam questions emphasize physiology more than spatial anatomical relationships. Imaging is logically not as well suited as gross dissection to learn physiological relationships of structures (such as muscle actions). Alternatively, there remains the possibility that radiographic instruction does positively impact written exam performance—as observed in the aggregate data—but was undetected because of the reduced sample size of the paired analysis.

The pragmatic findings from studying an in-class intervention are inevitably limited to some degree by uncontrollable personal characteristics of individual participants. Paired analysis of each student’s scores with and without using the study guides and normalizing each of the test scores controlled for important confounders. This is supported by the increase in explained variance in the paired analysis over the aggregate analysis. Nonetheless, there remain notable limitations to our study.

First, our analysis of study guide efficacy in the context of a live course strengthens its validity as a practical instrument but allows for some sources of variance beyond our control.

### TABLE 3. Distribution of Study Guide Completion and Corresponding Means of Normalized (Z) Scores

<table>
<thead>
<tr>
<th>Number of Guides Used</th>
<th>Number of Students n (%)</th>
<th>Mean ± SD of Normalized Radiology Practical Exam Scores</th>
<th>Mean ± SD of Normalized Gross Practical Exam Scores</th>
<th>Mean ± SD of Normalized Written Exam Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>21 (22.6) 0.08 ± 0.76</td>
<td>0.01 ± 0.67 -</td>
<td>0.00 ± 0.67 -</td>
<td>0.00 ± 0.66 -</td>
</tr>
<tr>
<td>1</td>
<td>8 (8.6) -0.06 ± 0.60</td>
<td>0.004 ± 0.57 0.34 ± 0.71</td>
<td>0.29 ± 0.53 0.46 ± 0.52</td>
<td>0.04 ± 0.64 0.10 ± 0.52</td>
</tr>
<tr>
<td>2</td>
<td>7 (7.5) -0.26 ± 0.95</td>
<td>-0.10 ± 0.69 0.12 ± 0.72</td>
<td>0.006 ± 0.54 0.14 ± 0.71</td>
<td>-0.07 ± 1.08 0.20 ± 0.62</td>
</tr>
<tr>
<td>3</td>
<td>18 (19.4) -0.42 ± 0.45</td>
<td>-0.33 ± 0.90 -0.17 ± 1.14</td>
<td>-0.25 ± 1.16 -0.17 ± 0.82</td>
<td>-0.44 ± 0.59 -</td>
</tr>
<tr>
<td>4</td>
<td>9 (9.7) -0.36 ± 0.68</td>
<td>0.03 ± 0.62 0.22 ± 0.46</td>
<td>0.17 ± 0.97* 0.20 ± 0.76</td>
<td>0.004 ± 0.94 0.20 ± 0.84*</td>
</tr>
<tr>
<td>5</td>
<td>8 (8.6) -0.61 ± 1.17</td>
<td>-0.71 ± 1.29 -0.07 ± 0.73</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>22 (23.7) -</td>
<td>0.40 ± 0.68 -</td>
<td>0.37 ± 0.76 -</td>
<td>0.44 ± 0.59 -</td>
</tr>
<tr>
<td>Total</td>
<td>93</td>
<td>-0.12 ± 0.99 0.20 ± 0.95*</td>
<td>-0.09 ± 0.92 0.17 ± 0.97*</td>
<td>0.004 ± 0.94 0.20 ± 0.84*</td>
</tr>
</tbody>
</table>

SD, standard deviation.

*P < .05.

*P ≤ .001, both for comparing scores with and without study guide use within each exam type by independent t-test.

![Figure 2](image-url)
that remain captured in the paired t-test analysis. For example, students may have personal preferences for certain body regions and adjust their studying accordingly. We must assume that changes within individuals are balanced throughout the diverse class (Table 2) and relatively small because a given student’s aptitude and study habits tend to be relatively stable attributes, especially over the course of only 2 months. Second, in drawing comparisons within each student’s performance, students who completed all or none of the study guides (22 and 21 students, respectively) could not be included in the paired analysis, thus excluding a notable portion of the class and limiting the study’s power. Finally, different student motivations for completing or not completing the guides cannot be fully reconciled. However, the similar distribution of module guide use by students who completed varying total numbers of guides suggests similar motivation to complete a given module. The bimodal distribution peaking at the first and fourth modules corresponds, for example, with the start of a new academic quarter after a period of recess. Additionally, students who were not motivated to complete many guides because they subjectively found them unhelpful or not time efficient would favor the null hypothesis.

In conclusion, our findings demonstrate quantitative support for the effectiveness of radiologic imaging to improve comprehension of anatomy, likely because of improved understanding of spatial anatomical relationships. Notably, the positive impact of radiologic instruction was observed with respect to both radiologic image interpretation and gross cadaver structure identification in independent exams. These results strongly support the need for further quantitative studies with larger sample sizes and utilization of various applications of radiology to better define the extent and methods by which radiology instruction can improve medical anatomy curriculum.

ACKNOWLEDGMENTS

We thank the students, faculty, and staff of the Pritzker School of Medicine, whose dedication to continuously improving medical education pedagogies made this study possible. We are also grateful to Jim O’Reilly, PhD, for instructional support, Cara V. Phillips for data compilation support, Charlene Sheridan for administrative support, and Kelly Ledbetter for graphics support.

REFERENCES