5-2012

Learning from graphically integrated 2D and 3D representations improves retention of neuroanatomy.

Farah Naaz
University of Louisville

LEARNING FROM GRAPHICALLY INTEGRATED 2D AND 3D
REPRESENTATIONS IMPROVES RETENTION OF NEUROANATOMY

By

Farah Naaz

B.Sc., University of Allahabad, 2004

M.Sc., University of Allahabad, 2006

A Dissertation

Submitted to the Faculty of the

College of Arts and Sciences of the University of Louisville

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

Department of Psychological and Brain Sciences

University of Louisville

Louisville, Kentucky

May 2012

LEARNING FROM GRAPHICALLY INTEGRATED 2D AND 3D
REPRESENTATIONS IMPROVES RETENTION OF NEUROANATOMY

By

Farah Naaz

B.Sc., University of Allahabad, 2004

M.Sc., University of Allahabad, 2006

A Dissertation Approved on

April 20, 2012

by the following Dissertation Committee:

John R. Pani, Ph.D

_____

Dissertation Director

Ronald D. Fell, Ph.D.

_____

Keith B. Lyle, Ph.D.

_____

Carolyn B. Mervis, Ph.D.

_____

Pavel Zahorik, Ph.D.

_____

DEDICATION

This dissertation is dedicated to my parents

Mr. Mohammed Mustafa Ansari

and

Mrs. Razia Begum

for encouraging me to follow my dreams.

ACKNOWLEDGMENTS

ABSTRACT

LEARNING FROM GRAPHICALLY INTEGRATED 2D AND 3D
REPRESENTATIONS IMPROVES RETENTION OF NEUROANATOMY

Farah Naaz

April 20, 2012

Visualizations in the form of computer-based learning environments are highly
encouraged in science education, especially for teaching spatial material. Some spatial
material, such as sectional neuroanatomy, is very challenging to learn. It involves
learning the two dimensional (2D) representations that are sampled from the three
dimensional (3D) object. In this study, a computer-based learning environment was used
to explore the hypothesis that learning sectional neuroanatomy from a graphically
integrated 2D and 3D representation will lead to better learning outcomes than learning
from a sequential presentation. The integrated representation explicitly demonstrates the
2D-3D transformation and should lead to effective learning.

This study was conducted using a computer graphical model of the human brain.
There were two learning groups: *Whole then Sections*, and *Integrated 2D3D*. Both groups
learned whole anatomy (3D neuroanatomy) before learning sectional anatomy (2D
neuroanatomy). The *Whole then Sections* group then learned sectional anatomy using 2D

representations only. The *Integrated 2D3D* group learned sectional anatomy from a graphically integrated 3D and 2D model. A set of tests for generalization of knowledge to interpreting biomedical images was conducted immediately after learning was completed. The order of presentation of the tests of generalization of knowledge was counterbalanced across participants to explore a secondary hypothesis of the study: *preparation for future learning*. If the computer-based instruction programs used in this study are effective tools for teaching anatomy, the participants should continue learning neuroanatomy with exposure to new representations. A test of long-term retention of sectional anatomy was conducted 4-8 weeks after learning was completed.

The *Integrated 2D3D* group was better than the *Whole then Sections* group in retaining knowledge of difficult instances of sectional anatomy after the retention interval. The benefit of learning from an integrated 2D3D representation suggests that there are some spatial transformations which are better retained if they are learned through an explicit demonstration. Participants also showed evidence of continued learning on the tests of generalization with the help of cues and practice, even without feedback. This finding suggests that the computer-based learning programs used in this study were good tools for instruction of neuroanatomy.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

Visualization is important in science education. Domains such as physics, chemistry, and biology rely heavily on visualization to explain concepts, processes, structures, and spatial relationships (Gilbert, 2007). Visualizations in the form of computer-based learning environments have been highly encouraged and are commonly adopted in science education (Miller, Lehman, & Koedinger, 1999; Lee, Linn, Varma, & Liu, 2010; Lowe, 2004; Russel & Kozma, 1994; Crowley & Medvedeva, 2006). However, most visualization tools are being developed on the basis of intuitions of the instructor or the designer and without taking findings in cognitive science into consideration (Ainsworth, 2006; Issenberg, McGaghie, Petrusa, Gordon, & Scalese, 2005; Khalil, Paas, Johnson, & Payer, 2005, Mikropoulos, & Natsis, 2011). There is a need for development of computer-based learning environments which are based on theories of learning and which have been rigorously evaluated for their effectiveness (Chariker, Naaz, & Pani, 2012; Cook & Levinson, 2011; Issenberg et. al, 2005; Khalil et al., 2005).

Neuroanatomy is a discipline that requires spatial visualization for a coherent understanding of the material. The brain has a complex spatial organization with a variety of different structures in complex relations to each other. Learning neuroanatomy is challenging due to the complexity of the brain and because the learner must assimilate a large amount of information from various anatomical representations, including atlases, cadaver dissection, and MRI images. Several recent studies have also pointed out that

1

students struggle in retaining their anatomy knowledge, and that there is a need for supplemental material to support the students in learning (Cottam, 1999; Fitzgerald, White, Tang, Maxwell-Armstrong & James, 2008, Waterson & Stewart, 2005).

Recently, many medical schools have started incorporating computer-based instructional programs to supplement anatomy education (Cottam, 1999; Dev, 1999; Dev, Hoffer, & Barnett, 2006). However, these computer-based teaching programs have typically not been evaluated for their effectiveness. A relatively small set of studies have focused on assessing the effectiveness of computer-based anatomical models in teaching anatomy. The results of these studies are mixed. Some studies report benefits of using computer-based programs over traditional material (Garg, Norman, & Sperotable, 2001; Luursema, Verwey, Kommers, & Annema, 2008; Luursema, Verwey, Kommers, Geelkerken, & Vos, 2006; Nicholson, Chalk, Funnell, & Daniel, 2006). Other studies report computer-based programs to be either as efficient as traditional teaching material (Codd & Chaudhury, 2011; Hariri, Rawn, Srivastava, Youngblood, & Ladd, 2004; Garg, Norman, Eva, Spero, & Sharan, 2002; Keedy, et al., 2011), or inferior to traditional approaches (Garg, Norman, Spero, & Maheshwari, 1999; Levinson, Weaver, Garside, McGinn, & Norman, 2007). However, the relatively poor performance by computer-based groups in these studies can be explained by several points. First, these studies utilize a short study time, which is unrealistic if the aim is to simulate a real-world learning environment. Second, these studies often teach relatively simple anatomy, such as carpal bone anatomy, which may not be spatially complex enough to require 3D visualization. Third, because building realistic and detailed computer graphical models

2

requires large amounts of resources, the computer-based models used in these studies are often simplistic, and less detailed than textbook illustrations.

Existing comparisons of computer-based instructional programs with traditional approaches to teaching do not allow adequate exploration of the capabilities that can be incorporated into computer-based programs. In order to develop a computer-based instructional program that can successfully facilitate the process of learning anatomy, there is a need to focus on designing programs which can provide realistic graphics, flexible learning environments, and that follow principles of learning described in cognitive science (Issenberg, et al., 2005; Silen, Wirell, Kvist, Nylander, & Smedby, 2008).

We have been developing computer-based instructional programs for teaching neuroanatomy and have evaluated its effectiveness empirically (Chariker, Naaz, & Pani, 2011, 2012; Pani, Chariker, & Naaz, in preparation). These studies focused on the challenges involved in learning sectional neuroanatomy. Sectional neuroanatomy is an important aspect of learning neuroanatomy, with application in domains such as radiology and pathology. Sectional neuroanatomy is the two dimensional (2D) representation of the brain through depictions, or actual samples, of planar slices taken from the three dimensional (3D) tissue at different orientations and positions. MRI images are the most familiar example, but students study many types of sectional representation (e.g., diagrams and dissections). The structures in the brain change their appearance in these sectional representations, in part because the whole brain has a complex spatial organization. For example, the fornix is a C-shaped structure which

3

wraps around the thalamus, runs through the hypothalamus, and connects the hippocampus with the mammillary body. The shape and location of the fornix changes in different sections of the brain. In addition, the fornix is confusable with the optic tract in some sections of the brain (but not others), as the structures sometimes appear similar. Thus, learning sectional anatomy becomes a very challenging task for the learner as it requires dealing with a large amount of complex spatial information. Graphical visualization tools should be helpful in understanding such complex spatial material. However, it is not clear what approach to visualization would best lead to a coherent understanding in this case. In these studies, different approaches to graphical visualization was explored to determine which approaches lead to better learning outcomes for sectional neuroanatomy.

A previous research on learning sectional anatomy focused on encouraging integration of the 3D and 2D representations of neuroanatomy (Chariker et al., 2011, 2012). The hypothesis was that learning whole anatomy (3D shape and organization of the neuroanatomical structures) before learning sectional anatomy would make learning sectional anatomy more efficient and effective. This hypothesis was supported by research on learning and memory which encourages presentation of information in an organized framework (Bower, Clark, Lesgold, & Winzenz, 1969; Bransford, Brown, & Cocking, 2000). A 3D computer graphical model was created to teach whole anatomy, and 2D graphical sections were created by dense sectioning of the 3D model to teach sectional anatomy. The performance of a group who learned whole anatomy first and then sectional anatomy (*Whole then Sections*) was compared with a group that learned only sectional anatomy (*Sections Only*). The results of the study showed that learning

4

whole anatomy before learning sectional anatomy was better than learning just sectional anatomy. The *Whole then Sections* group was more efficient in learning neuroanatomy. They learned both whole and sectional anatomy in little more time than the *Sections Only* group, while making significantly fewer errors. In addition, the *Whole then Sections* group retained their sectional anatomy knowledge better than the *Sections Only* group after an interval of 2-4 weeks. Another important benefit of learning whole anatomy before sectional anatomy was found in the generalization of knowledge. The tests of generalization of knowledge were conducted using biomedical images, which participants had never seen before, to test their capacity to transfer their knowledge to new representations of sectional anatomy (MRI and Visible Human images). The *Whole then Sections* group was significantly better at generalizing their knowledge to new representations of sectional neuroanatomy than the *Sections Only* group. These results suggest that whole anatomy knowledge provided a framework for organization of sectional anatomy knowledge, hence leading to more efficient learning, better retention, and better generalization of sectional anatomy knowledge.

A second study attempted improving the organization of sectional anatomy knowledge by interleaving (alternating) trials of whole and sectional anatomy learning (Pani et al., in preparation). This approach to learning provides multiple opportunities for the participants to transfer their whole anatomy knowledge to learning sectional anatomy, and sectional anatomy knowledge to learning whole anatomy. Evidence from the literature on interleaved learning suggests that although learning from an interleaved presentation is more difficult than learning from blocked presentation, it leads to better learning outcomes (Kornell & Bjork, 2008; Rohrer & Taylor, 2007; Taylor & Rohrer,

2010). Learning from a sequential (blocked) presentation of whole and sectional anatomy (*Whole then Sections*) was compared with learning from an interleaved presentation of whole and sectional anatomy (*Alternation*). Neuroanatomy was taught from three different views (i.e., front, side, and top). The presentation of the views was blocked. Participants learned one view to the criterion level, then started learning the second view, and so on. It was found that interleaved presentation makes learning more efficient than learning from a sequential presentation. The *Alternation* group learned sectional anatomy with fewer errors than the *Whole then Sections* group. The benefit of interleaved presentation was also seen in the number of trials to finish learning a view. The *Alternation* group took significantly fewer trials to learn sectional anatomy in the third (and last) view of sectional anatomy. The benefit of interleaving did not extend to long-term retention and generalization of knowledge. Both the groups performed at comparable levels on these tests.

Although both of these approaches to learning neuroanatomy (sequential and interleaved) were efficient ways to learn, there were some brain structures which remained challenging for the participants. The long-term retention tests of sectional anatomy showed that while most test items were retained at a high level (group retention of 80-90%), retention for some test items was quite low (40% or below). These test items appeared to be difficult for participants because they were confusable with each other due to their common size, shape, and location (Chariker et al., 2012). It is possible that learning and retaining these difficult test-items would be easier if participants understood how the 3D to 2D transformation of structures occurs in sectional representation.

Based on the previous two studies, it appears that integrating 3D and 2D information may be a key to a coherent understanding of 3D and 2D mapping. The sequential and interleaved presentations of whole and sectional neuroanatomy relies on participants' ability to recall and integrate the 3D and 2D representations. The literature on spatial cognition shows that spatial knowledge is often global, and imagining even simple spatial transformations depends on favorable conditions (Hinton, 1979; Kozhevnikov, & Hegarty, 2001; Pani, Jeffres, Shippey, & Schwartz, 1996; Pani, Chariker, Dawson, & Johnson, 2005; Reed, 1974; Stevens & Coupe, 1978; Tversky, 1981). For example, Tversky (1981) showed that memory for spatial information, such as the location of continents on a map, is global and schematic. Due to tendency for spatial alignment, people remember North and South America as vertically aligned when actually they are not. Another interesting example is of imagining projective transformations (e.g., casting a shadow of an object). People struggle with imagining the shadow of even simple objects such as a cube if the object is not shown aligned to the direction of projection (Pani et al, 1997). These studies suggests that many instances of spatial integration of whole and sectional representation can be too challenging for the learner to attain without seeing them explicitly. Certainly there are cases when the best way to teach a spatial transformation is to show it to people (Pani et al., 2005). Hence, instead of relying on the learners to integrate the information on their own, it may be a better approach to show them the explicit mapping of whole and sectional anatomy in an integrated graphical representation. If seeing the explicit 2D-3D spatial transformation is advantageous for better understanding of the difficult test items, performance on these test items should improve after learning with explicit graphical integration. Therefore, in

7

this study learning neuroanatomy from an explicitly integrated graphical representation was compared with sequential presentation of neuroanatomy to explore if there is any benefit of learning from an integrated representation.

While the literature on spatial cognition suggests that explicit integration of spatial information may be beneficial, an alternative hypothesis is that cognitive integration of whole and sectional anatomy is always a better approach for integration. The effort required to integrate the representations cognitively may be an instance of desirable difficulty (Bjork, 1994). The concept of desirable difficulty suggests that conditions which may seem difficult during instruction (such as spaced learning and interleaved presentation) may lead to better retention than conditions which appear to be more efficient. Providing a graphical tool for a process that is best performed cognitively may impair learning. There is evidence that complex graphical visualizations are not always beneficial (Hegarty, 2004; Keehner, Hegarty, Cohen, Khooshabeh, & Montello, 2008; Tversky, Morrison, & Betrancourt, 2002). In a previous study (Chariker et al., 2011), providing participants the capability for continuous navigation between sections during sectional anatomy learning did not lead to any benefit over discontinuous navigation (similar to turning pages in an atlas). It is possible that learning from an explicitly integrated representation would lead to poor retention of sectional anatomy. However, we believed that neuroanatomy is complex enough that learning from an explicitly integrated graphical model of 2D and 3D neuroanatomy would help the learner to build a more accurate and organized mental model of sectional anatomy. This, in turn, would benefit learning and retention. Note that if participants in the study show better learning with an explicitly integrated graphical model of 2D and 3D anatomy, it would be

8

one instance where computer-based learning provides a learning opportunity which is otherwise very difficult to obtain.

A major concern for any approach to learning is whether it is suitable for learners with varying cognitive abilities. Some studies on visualization suggest that visualizations may only be beneficial for high spatial learners (Cohen & Hegarty, 2007; Garg et al., 1999; Keehner et al., 2008; Hoffler, 2010; Huk, 2006; Levinson et al.,2007; Nguyen, Nelson, & Wilson, 2012). It is important to note that most of the studies that report no benefit for low spatial groups were focused on the initial stage of learning after a brief exposure to the learning material. These studies on spatial learning suggests that there will be large variations in how beneficial visualizations can be to learners of varying spatial ability during the initial learning stage. However, these differences may not be important if the goal is mastery of the material. Studies using longitudinal learning suggest that with practice, even low spatial groups can achieve proficiency (Chariker, 2009; Keehner, Lippa, Montello, Tendick, Hegarty, 2006; Pani et al., 2005).

Based on the studies of visualization that report no benefit for the low spatial group, it could be predicted that low spatial participants in this study will not benefit from graphically integrated visualizations. On the other hand, studies using longitudinal learning suggest that graphically integrated visualizations may be helpful for low spatial participants (Chariker, 2009; Keehner et al., 2006). Since learners in this study were allowed to learn until they reached a high performance criterion, even low spatial participants should benefit from the integrated visualization tools. Overall, the role of spatial ability in learning from graphical programs remains an empirical question.

The procedure for testing generalization of knowledge was modified in this study to address a secondary hypothesis about *preparation for future learning* and the present methods of computer-based learning. The concept of *preparation for future learning* suggests that the benefits of learning in real-world situations go beyond immediate outcomes (Bransford & Schwartz, 1999). Another way to evaluate the effectiveness of an approach to learning is to measure how prepared the learner is to continue learning in the field. Participants in this study were tested on their ability to generalize knowledge to new representations of neuroanatomy such as MRI images and photographs of real brain sections (Visible Human images). These generalization tests were used for two purposes. The initial purpose was to evaluate between-group differences in performance on the ability to generalize knowledge of neuroanatomy. The second purpose was to evaluate if participants continued learning with exposure to new representations of neuroanatomy in the context of cues, even in the absence of feedback. For example, if a cue is provided to participants in the form of an arrow pointing to a structure, and their task is to name the structure, that implies that the participants have been taught that structure. Biomedical images contain much more information than the participants have seen during learning. If the participants remembered their neuroanatomy well, the cue in the form of an arrow indicating a structure would help participants eliminate other neuroanatomical structures that they might consider as possible locations for that structure. In later exposure to biomedical images with different types of cues, the previous experience with the material should provide some help in correct identification of the structure. Thus, learning can continue with new representations of neuroanatomy even in the absence of feedback. However, the initial learning had to be relatively complete and generalizable.

There were two learning groups in the study: *Whole then Sections* and *Integrated 2D3D*. Both groups learned whole anatomy first, because it is efficient to learn with the present methods, and it provides a good overview of neuroanatomy. Both groups then moved to learning sectional anatomy, either with sectional representation only or with a graphical integration of whole and sectional anatomy.

Both the learning groups used a method of learning called adaptive exploration (Pani et al., 2005; Chariker et al., 2011, 2012; Pani et al., in preparation). Adaptive exploration provides flexibility to the learner in graphically exploring the learning material. This capability is embedded in learning trials which consist of successive stages of study, test, and feedback. As a consequence, participants space their learning, frequently test their knowledge, and receive continuous feedback. This approach was adopted because there is evidence to suggest that these methods lead to good retention of information (Melton, 1970; Karpicke & Roediger, 2008; Kornell & Bjork, 2009). Another important aspect of this study was that the participants continued learning the material until they reached a high performance criterion. Requiring a high performance criterion allowed us to test the participants' long-term retention and generalization of knowledge after participants had fully learned the material.

METHOD

Participants

Sixty-four volunteers were recruited for the study through paper and online advertisements posted throughout the University of Louisville campus. Eleven participants did not complete the study due to scheduling conflicts or because they did not comply with instructions. Participants were randomly assigned to the learning groups while keeping the group mean and distribution for DAT score balanced. A further balancing procedure was used to match the participants in the two learning groups on a pre-test of sectional neuroanatomy identification (see Procedure). After matching participants in the two learning groups, forty participants were included in the final analyses. The participants included in the final analyses were between 16 and 34 years of age.

All volunteers took an initial survey which assessed their knowledge of neuroanatomy. The survey asked them to indicate if they were familiar with the names of the 19 neuroanatomical structures included in the study and whether or not they would be able to locate these structures in the brain. Only the volunteers with minimal or no knowledge of neuroanatomy were included in the study.

Participation in the study required coming to the laboratory for an hour long session at least twice a week. On average, participants took twelve sessions to finish. Participants were compensated 100 dollars for completing the study. Participants who did not complete the study were compensated at the rate of 8 dollars per hour, up to a total of

100 dollars. This rule for compensation was implemented to discourage any intentional delay in completion of the study.

Material

  Psychometric tests. Before learning neuroanatomy, participants were tested on their spatial ability and the capacity of their short-term memory for visuospatial patterns using standard psychometric tests. The Differential Aptitude Test: Space Relations (DAT; Bennett, Seashore, & Wesman, 1989) was used to measure spatial ability. The DAT is a test of "mental paper folding". The test involves presenting a two-dimensional pattern which can be folded to form a three-dimensional figure. There are four pictures of three-dimensional figures presented as possible answers. The participant's task is to indicate the correct option among the four provided. There were thirty-five test questions. Percentile scores for the test were used in the data analyses.

  The Designs I test for immediate recall from Wechsler Memory Scale-IV (Wechsler, 2009) was used to measure short-term memory for visuospatial patterns. The test consists of separate measures for visual content memory, visual spatial memory, and an overall visuospatial memory score. The participants were shown a picture of a 4x4 grid. The grid contained patterns in some of its cells. The participants' task was to remember these patterns and their location in the grid. After the picture was removed, participants were presented with a physical 4x4 grid with empty cells and a set of cards. These cards included both the correct patterns and distractors. Participants' task was to identify the cards with the correct patterns and to place them in the correct locations on the grid.

The visual content memory score (Visual Content) represent the participants'

ability to identify the correct patterns irrespective of the spatial location in the grid. The

spatial memory score (Spatial Memory) represent the participants' ability to remember

the correct location on the grid irrespective of the accuracy in choosing the correct

pattern. The overall visuospatial memory score (Overall Memory) was calculated using

the sum of content and spatial memory score plus bonus points for remembering both the

correct patterns and the locations. The test consisted of four presentations with increasing

levels of difficulty. The first item showed 4 patterns in the grid, the second and third

items showed 6 patterns in the grid, and the fourth item showed 8 patterns in the grid.

The scaled scores for Visual Content, Spatial Memory, and Overall Memory were used in

the study. Scaled scores range from 1-19, with a mean at 10 and standard deviation of 3.

These scores represent the participant's performance relative to their peers in the

reference age group.

Neuroanatomical model. A 3D computer graphical model of the human brain

was adopted from earlier studies (Chariker et al., 2011; Pani, et al., 2012) This model was

created using digital photographs of the cryosections of a cadaver brain available from

the Visible Human project (version 2.0) of the National Library of Medicine (Ackerman,

1995; Ratiu, Hillen, Glaser, & Jenkins, 2003). The model consisted of 19 major

neuroanatomical structures: amygdala, brainstem, caudate, cerebellum, cortex, fornix,

globus pallidus, hippocampus, hypothalamus, mammillary bodies, nucleus accumbens,

optic tract, pituitary, putamen, red nucleus, subthalamic nucleus, substantia nigra,

thalamus, and ventricles. Dense sectional representation of the 3D model also was

available in the three standard planes: coronal (frontal), sagittal (side), and axial (top), for

14

use in the sectional anatomy learning programs. Neuroanatomy was taught in the study from all three standard views.

Learning programs. There were three learning programs in this study: one program for learning whole anatomy and two programs for learning sectional anatomy. All learning programs were used in series of individual learning trials. Each learning trial was a continuous sequence that consisted of a timed study stage (4 minutes), a self-timed test stage, and a timed feedback stage (2 minutes).

The three learning programs differed in the type of neuroanatomical model (whole or sectional) being used, and the graphical interaction, but the basic structure of trials was the same across all three learning programs. During study, participants could freely explore the brain model using the graphical tools available. Clicking on a structure with a computer mouse highlighted the structure, and its name appeared prominently at the bottom of the screen. In the test stage, participants were asked to name the structures they had learned by clicking on each structure and selecting its name from a button panel. The feedback stage consisted of two elements: a numerical feedback screen, and graphical feedback. The numerical feedback screen was presented first, and it indicated the number of structures that were named correctly, the number of structures named incorrectly, and the number of structures omitted. The graphical feedback consisted of a color coding of the brain model: the structures named correctly appeared in green, the structures named incorrectly appeared in red, and the structures that were omitted appeared in their original color. Participants could click on structures, highlight them, and see their names. The feedback stage ended with the appearance of an exit screen which displayed the participant's percentage correct score on the test.

Whole anatomy program. The program for learning whole anatomy featured the 3D model of the brain. Both of the learning groups (*Whole then Sections, Integrated 2D3D*) learned whole anatomy using this program. Participants could explore the model freely with the help of tools provided for rotation, zooming, and virtual dissection (Figure 1). Participants could select a structure and learn its name. The brain model initially appeared from a coronal, sagittal, or axial view, and participants could rotate the brain by 360 degrees in the horizontal plane of the scene and a vertical plane of the brain. The virtual dissection of the brain was conducted by selecting structures and removing them one by one, using buttons on the button panel at the right side of the screen. Participants also could restore the structures removed either one by one or all at once by using the buttons on the button panel.

The test stage interface was similar to the study stage, with the exception that the model was restricted to the view (coronal, sagittal, or axial) in which it was originally presented in the study stage (Figure 2). Participants could rotate the brain only 45 degrees from the canonical view. Tools for zooming, removing, and restoring the structures were also available. To name a structure, participants selected the structure, selected its name from the button panel, and then clicked on a "Submit Name" button. Structures which were named turned blue to help the participants keep track of the structures they had already named. A numerical counter also was included at the top-left corner of the screen to indicate the number of structures named and the number of structures remaining. Participants could omit naming a structure if they wished. Participants exited the test stage once they finished naming the structures by clicking on the "Proceed to Feedback" button. The participants received numeric feedback first and then proceeded to graphical

feedback by clicking on a button. The participants could explore the color-coded brain

model using all the tools that were available to them during the study stage (Figure 3).
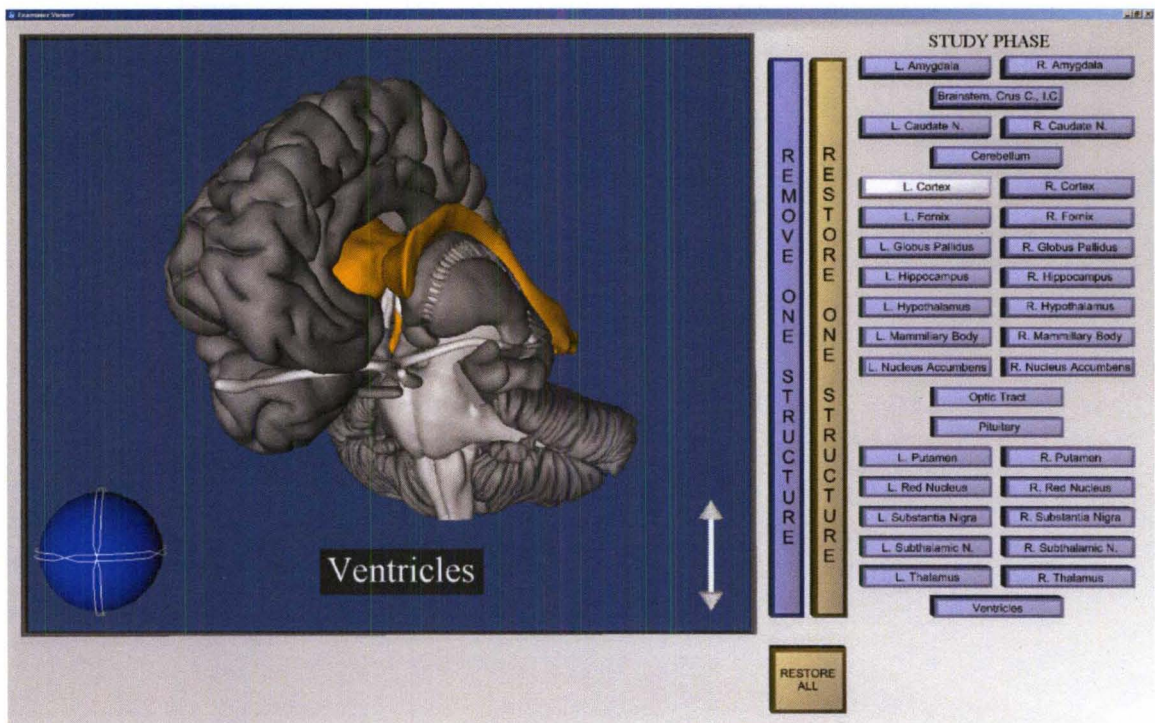
Figure 1. Whole anatomy program: Study stage. The left cortex in the model has been removed and the ventricles have been selected in this image.

Structures Submitted: 3
Structures Omitted: 16

REMOVE ONE STRUCTURE

TEST PHASE

Amygdala
Brainstem, Crus C., I.C.
Caudate N.
Cerebellum
Cortex
Fornix
Globus Pallidus
Hippocampus
Hypothalamus
Mammillary Body
Nucleus Accumbens
Optic Tract
Pituitary
Putamen
Red Nucleus
Substantia Nigra
Subthalamic N.
Thalamus
Ventricles

RESTORE ALL

SUBMIT ANSWER

PROCEED TO FEEDBACK

Figure 2. Whole anatomy program: Test stage. The left cortex has been removed to name inner structures. The structures in blue have been named. Structures in their original color have not been named.
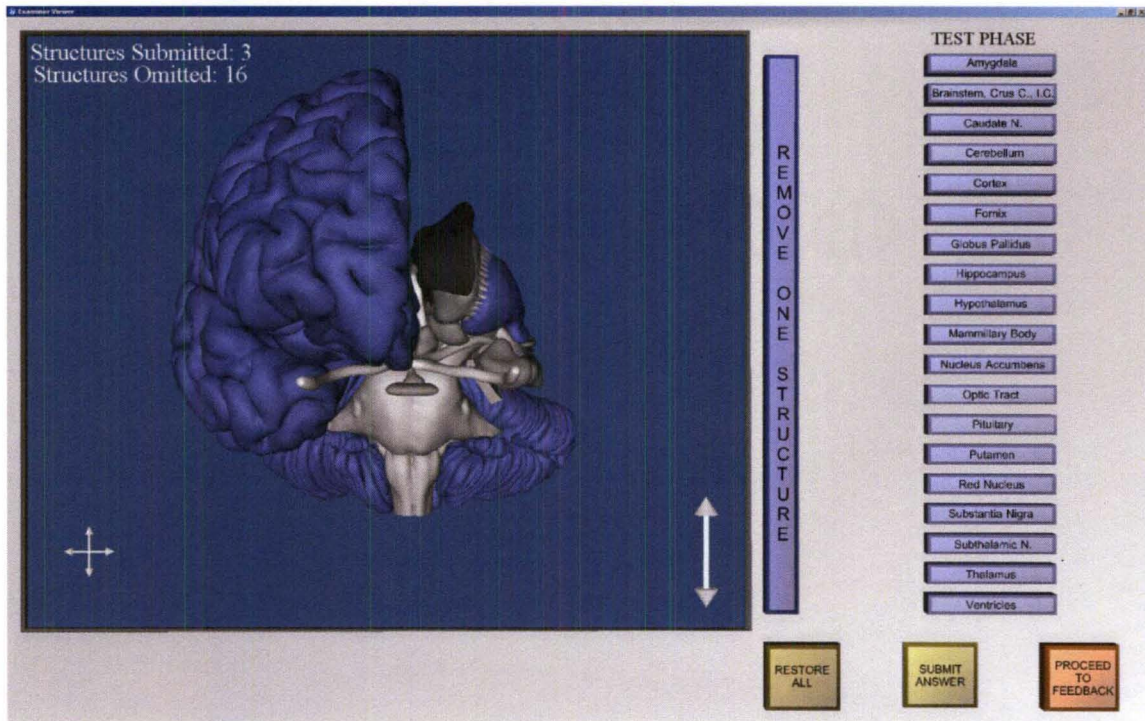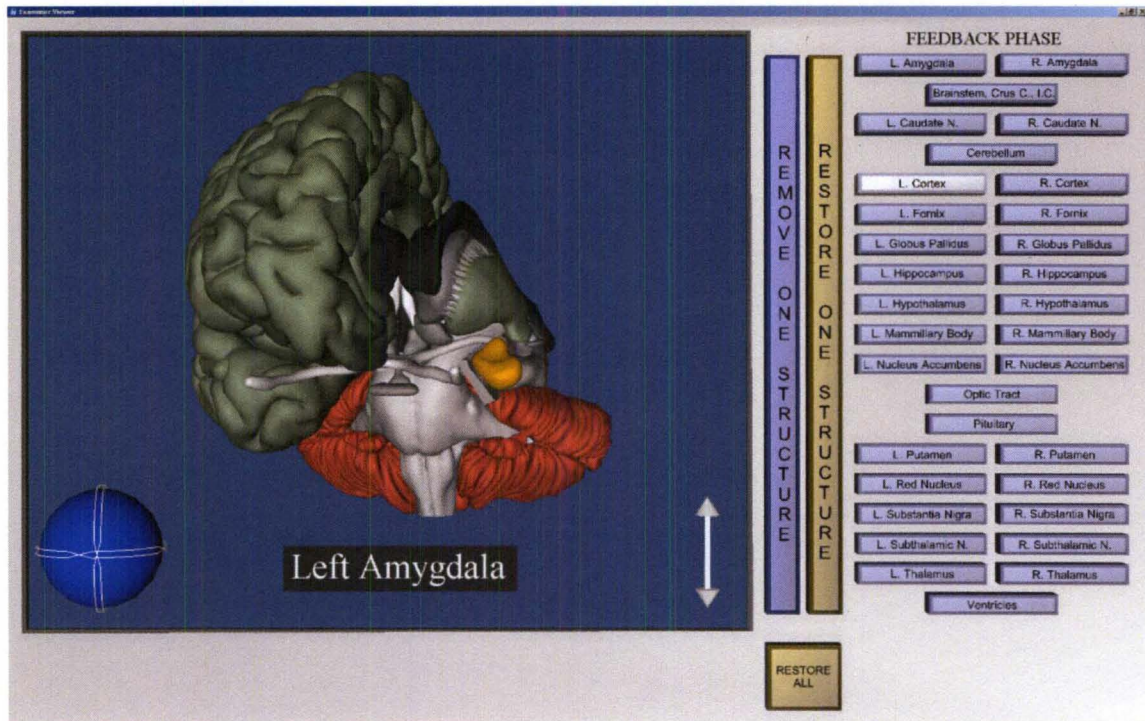
Figure 3. Whole anatomy program: Feedback stage. The left cortex has been removed. The structures that have been named correctly appear green in the model. The structures that have been named incorrectly appear red, and the structures which were omitted appear in their original color.

Sections only program. The *Whole then Sections* group learned sectional anatomy using a program that permitted exploring the 2D model of the brain (called the Sections Only program). During study, the participants saw thin parallel sections of the brain in a single view (coronal, sagittal, or axial). The participants could move from one section to another by using a slider tool provided at the bottom of the screen (Figure 4). For example, in the coronal view, moving the slider from the left end to the right allowed participants to see all the sections of the brain from front to back. Participants could go to individual sections of the brain and click on structures to learn their names. Clicking on a structure highlighted it, and its name appeared at the bottom of the screen.

The sectional anatomy test stage was slightly different from the whole anatomy test stage. Here only a few structures were selected from some of the sections for testing (Figure 5). This procedure was adopted because all sections of the model could not be tested in every trial, and specifying the test items allowed us to vary them in different trials and to adequately sample each structure of the brain. Several sections of the model (12-15) were selected for each test stage. The structures that were tested were indicated by arrows in the sections. The participants were to name the structures indicated by the arrows. Once again, participants named a structure by selecting it, selecting its name from the button panel, and clicking the "Submit Answer" button. As with whole anatomy testing, tested structures turned blue, and a counter was provided to keep track of the number of structures named in a test section. In addition, a counter was provided at the bottom of the screen indicating the number of test sections remaining. Participants could move from one test section to the next by clicking on the "Next Section" button, but they could not go back to a previous test section.

After finishing the test stage, participants received numeric feedback followed by graphical feedback. The graphical feedback stage was similar to the study stage and included all the sections of the brain (Figure 6). The test sections were marked "Test Section". The arrows indicating test items remained, and the test items again were color coded to reflect the performance on the test. The feedback stage allowed participants to go freely through all of the sections and to see where the test sections were taken from the model. They could select structures and see their names. The trial concluded with the appearance of the exit screen.
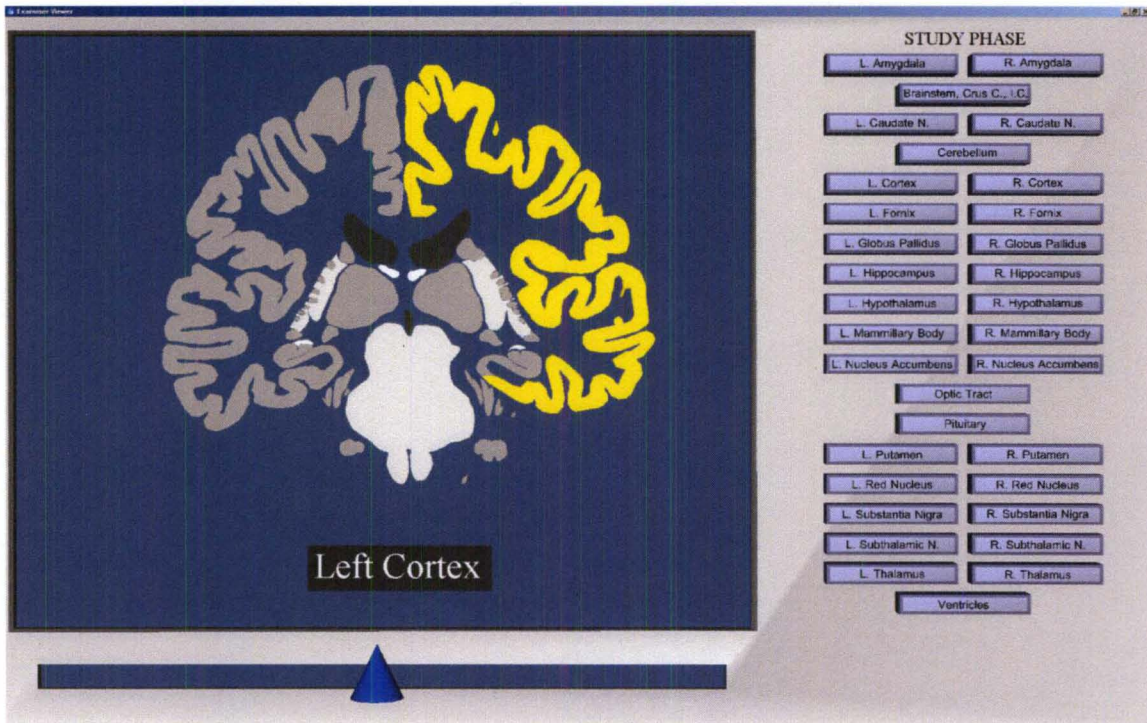
Figure 4. Sections only program: Study stage. The left cortex has been selected in a coronal section from the middle of the brain.
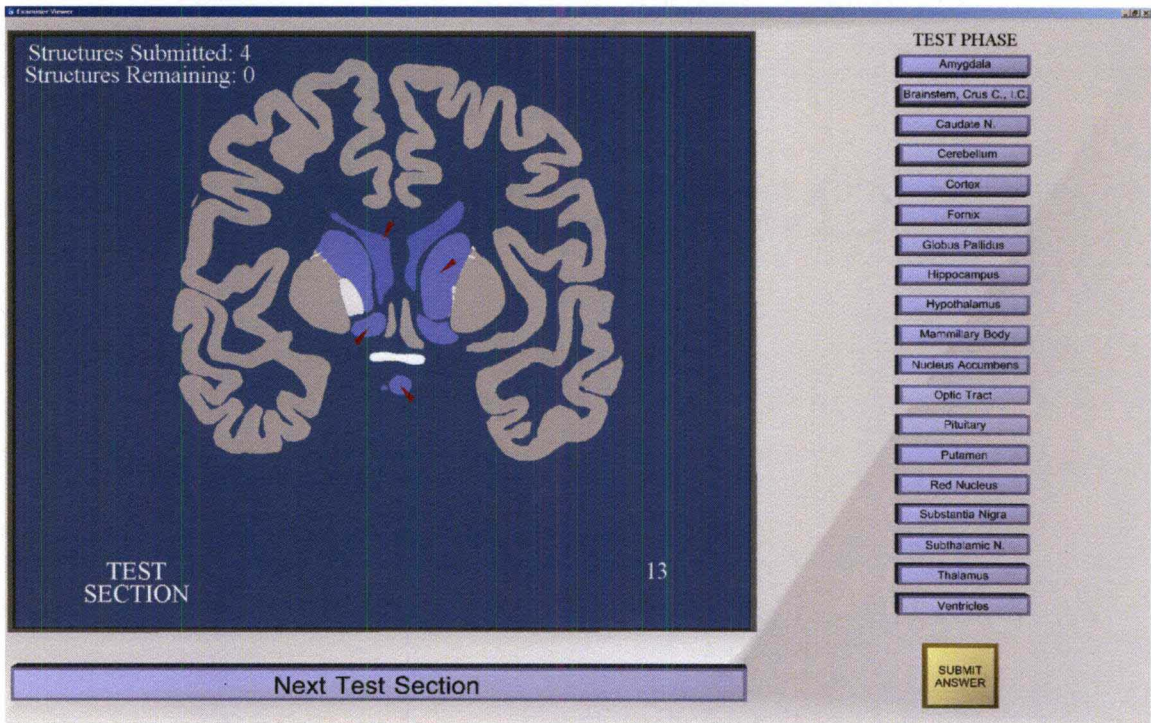
Figure 5. Sections only program: Test stage. Red arrows indicated structures to be named in each test section. The structures that have been named appear in blue.
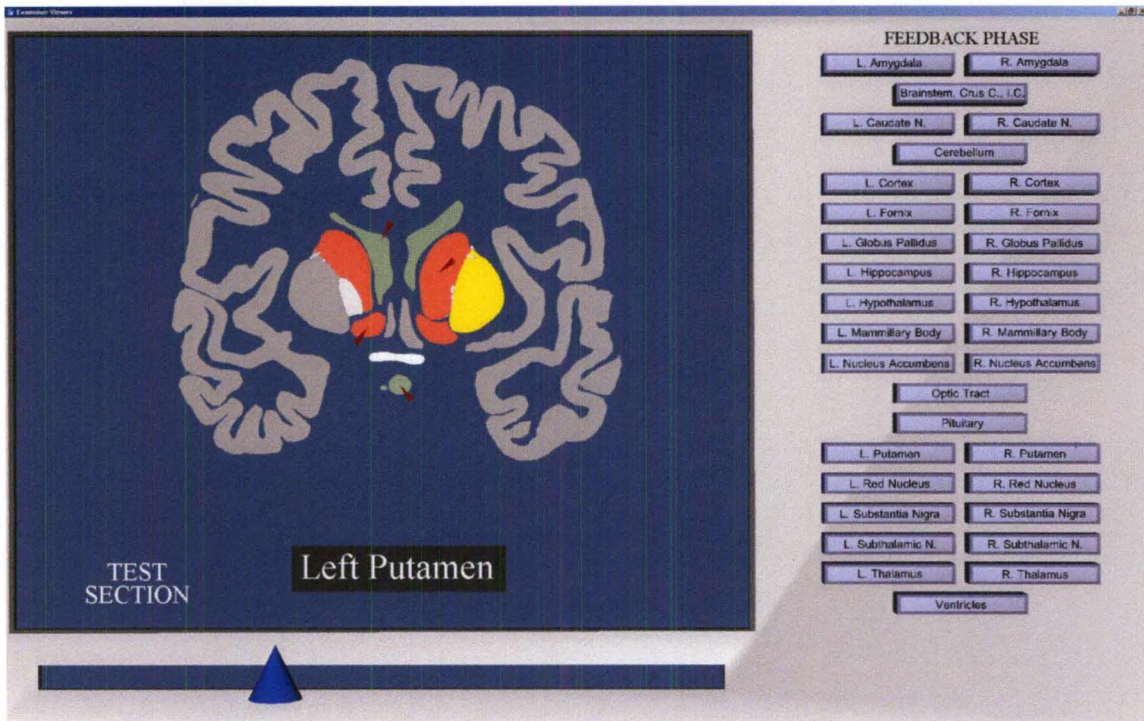
Figure 6. Sections only program: Feedback stage. Test sections appeared with their red arrows. The structures that have been named correctly appear green in the model. The structure that have been named incorrectly appear red, and the structures which were omitted appear in their original color.

2D3D program. The *Integrated 2D3D* group learned sectional anatomy using the

2D3D program (Figure 7). During the study stage of a trial, participants could explore the

3D model using all the tools they used earlier in the whole anatomy program for rotation,

zooming, and virtual dissection. In addition, the 3D model could be sliced to view

sectional anatomy with the help of the slider tool. On moving the slider from left to right,

a virtual cutting plane appeared to move through the brain and to remove the part of the

brain in front of the plane. This exposed the 2D cross-sections of the 3D structure from

one end of the brain to the other. For example, in the coronal view moving the slider from

left to right made a cutting plane move from the front of the brain to the back, removing

the part of the brain anterior to the plane. Beyond the cutting plane, the 3D structures

remained visible, allowing the participants to see the transformation from 3D to 2D.

Participants could move the slider back and forth to see how these 3D structures would

appear in sections from different depths in the brain. The test and feedback stages were

identical to the test and feedback stages of the sections only program (i.e., test and

feedback included only the 2D sectional anatomy, see Figures 5 and 6).

Figure 7. 2D3D program: Study stage. The program allowed exploration of whole anatomy and sectional anatomy simultaneously, using an integrated graphical representation. In this screenshot the left cortex is removed to expose the inner structures. The slider tool was used to remove the sections from the front of the brain.
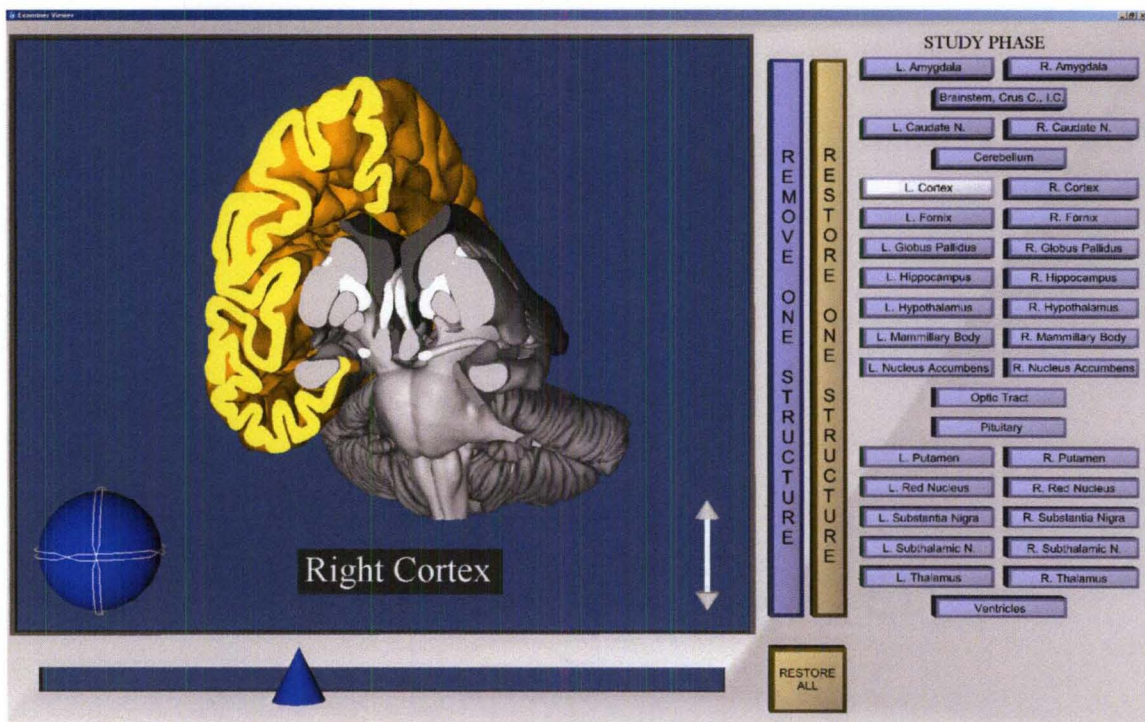
Generalization tests. Four generalization tests were created using biomedical images to test the transfer of knowledge to new representations of neuroanatomy. The biomedical images used were MRI images (from the SPL-PNL Brain Atlas, Kikinis et al., 1996), and images from the Visible Human project. The Visible Human (VH) images were photographs of cryosections of a human cadaver through the brain and skull. Compared to the VH images, the MRI images had low resolution, low contrast, and few clear boundaries between structures. Both set of images showed many additional structures, such as bones, muscles, and blood vessels, that were not taught during learning. Both the VH images and the MRI images were presented from all three views.

Global cues test. In this test, the participants were presented with a set of images one by one, and they were asked to name all recognizable structures in every image. The participants received two global cues to orient them for this test, one visual and one numeric (Figure 8). The visual cue was a brain icon at the top left corner of the screen which showed a line going through the brain, indicating the orientation and depth of the cross-section. For example, in the coronal view, a brain icon was presented from a side view with a line indicating the position and orientation of the coronal slice. The numeric cue was presented below the brain icon indicating the number of identifiable structures in that image. Participants indicated a structure by clicking on the image where they thought the structure was present and a dot would appear on the image at that location. They then selected a corresponding name from the button panel for the structure and submitted the answer. Participants were asked to give their best effort on a particular test image and then move to the next image.

Figure 8. Generalization test: Global cues. The participants were asked to identify all recognizable structures in the biomedical images. The brain icon provided information about the orientation and the position of the image. The numeric cue provided information about the number of recognizable structures present in the image.

Submit structure. In this test, participants were presented with a set of images. The name of a structure appeared at the bottom of the screen for each image (Figure 9). The task was to indicate the location of the structure being named in the image. Participants responded by clicking on the image where they thought the structure was located, and a dot appeared on the screen at that location. The participants then submitted their response and moved to the next test image.
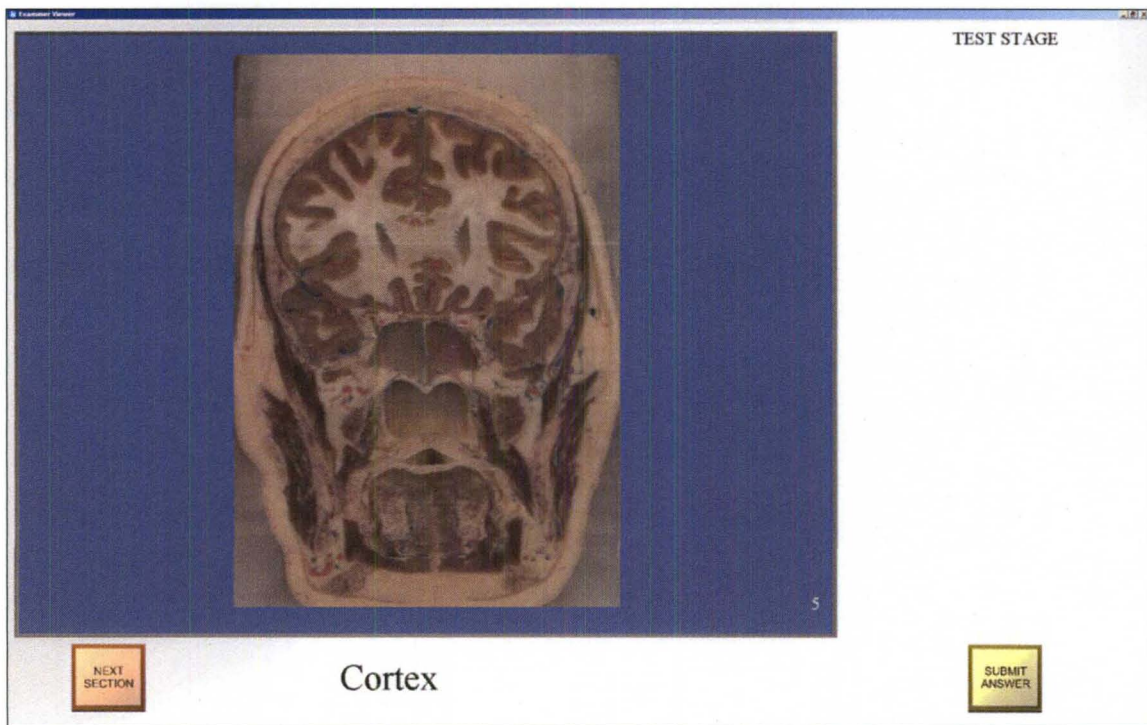
Figure 9. Generalization test: Submit structure. The participants were asked to indicate the location of the structure named at the bottom of the screen.

Submit name. In this test, participants were presented with a set of images. A single structure was indicated by a red arrow on each image (Figure 10). The participant named the indicated structure by selecting a name from the button panel and submitting the answer.
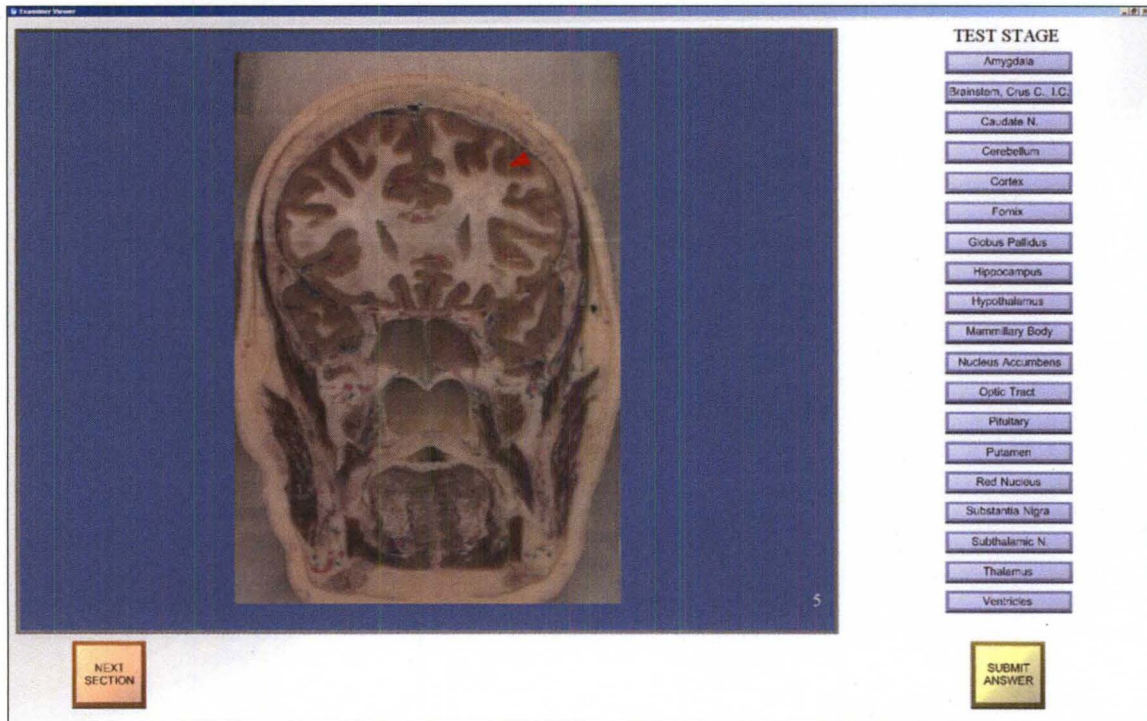
Figure 10. Generalization test: Submit name. The participants were asked to name the structure being indicated by the red arrow.

Numeric cues. This test was similar to the Global Cues test but included only the numeric cue about the number of structures present in the image (Figure 11). Participants were instructed to name as many structures as they could in each image by clicking on the structures and using the button panel to name them.

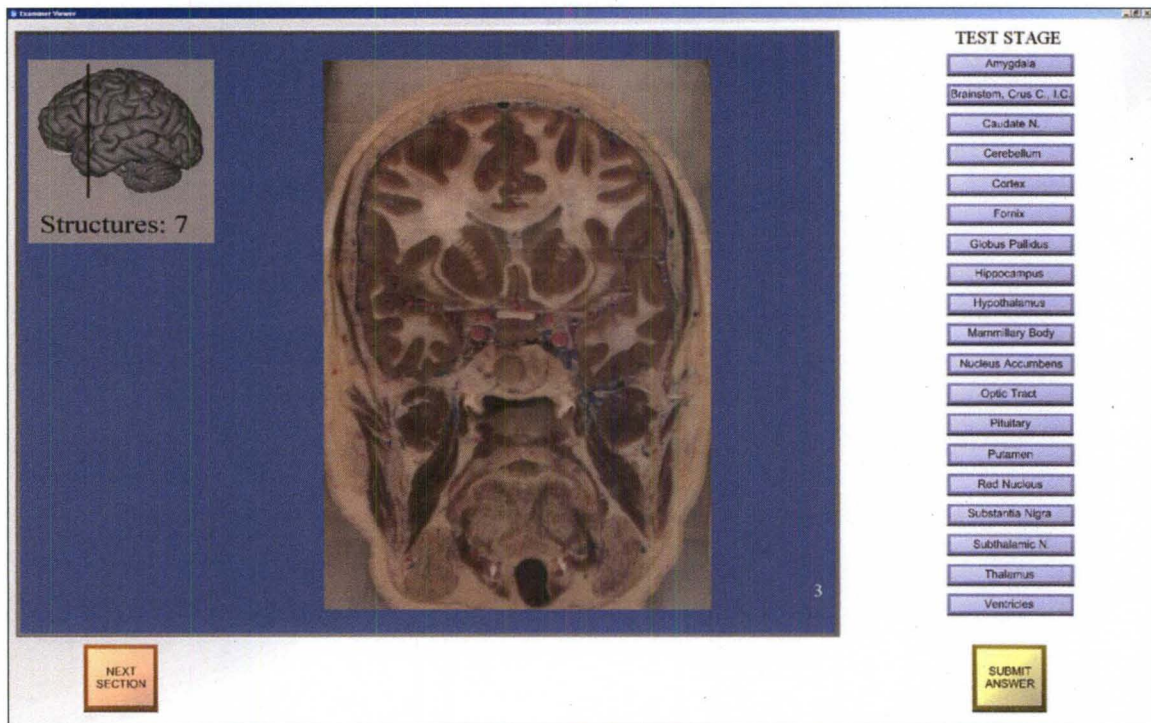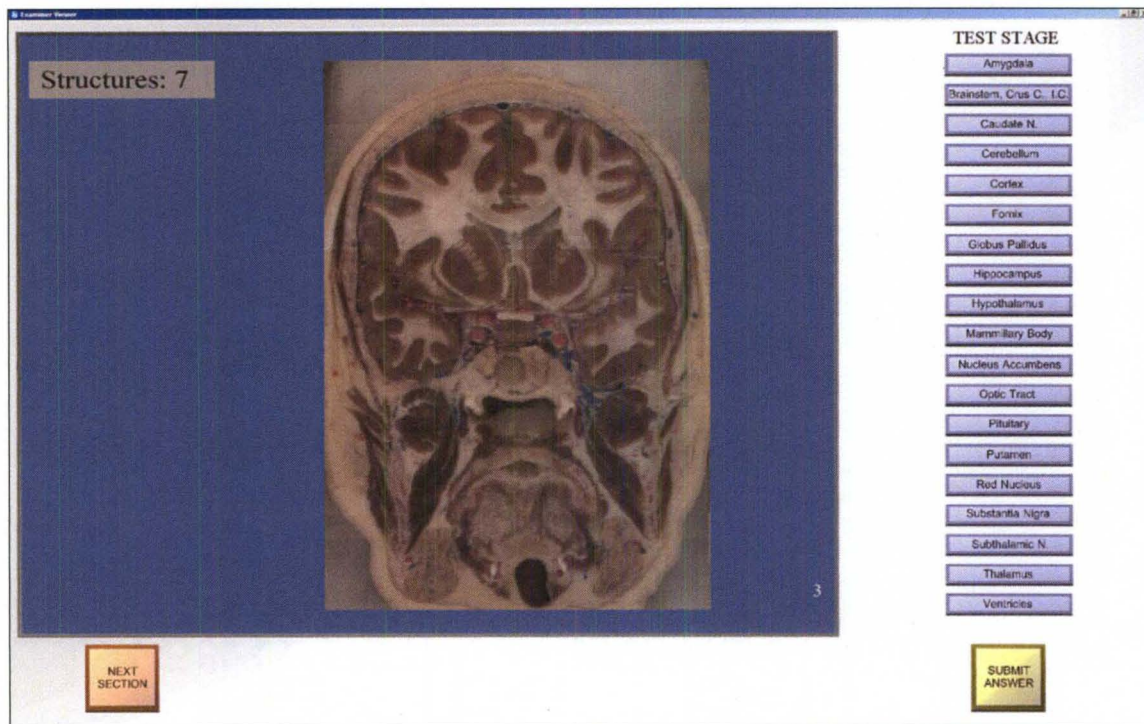*Figure 11*. Generalization test: Numeric cues. The participants were asked to identify all recognizable structures in the biomedical images. The numeric cue provided information about the number of recognizable structures present in the image.

Sections test and long-term retention test. The Sections Test and Long-term Retention Test (LTR) were tests of knowledge of sectional anatomy. These tests were taken from the tests used in the sectional anatomy learning trials. Hence, they were identical to the test stage of sectional anatomy trials. Thus, several sections of the 2D model were presented to the participants. They had to name the structures that were indicated by arrows by clicking on the structures and selecting the names from the button

panel. Counters were provided to help keep track of the number of structures named, the number of structures remaining in the section, and the number of test sections remaining in the test. No feedback was provided after these tests.

Instruction programs. Instruction programs were created for all learning programs and generalization tests. A mock brain model was created for instruction of the learning programs (Whole Anatomy, Sections Only, and 2D3D). The mock brain model contained several geometric structures in a variety of shapes and spatial relations. The structures were given pseudo-biological names. Using the mock brain model in the learning programs, demonstration videos were created which included visual demonstration and verbal description of how to use the learning programs. After watching the videos, the participants demonstrated their understanding of how to use the program by using the instruction programs with the mock brain. The experiment did not proceed until the participant could successfully demonstrate complete use of the relevant programs.

The instruction programs for the generalization tests consisted of biomedical images (MRI and VH images) that were not part of the actual tests. Participants received instruction on how to use the generalization test programs.

Apparatus

The study was conducted using individual workstations with high capacity graphics cards and sufficient RAM for smooth presentation of the neuroanatomical models and images. The programs were displayed on twenty-four inch LCD monitors at a resolution of 1920 x 1200 pixels. The participants were seated approximately 60 cm from the monitors. Participants learned by themselves in quiet rooms using the

neuroanatomical programs. They were provided with headphones to be used for the instructional videos.

Design and Procedure

The primary hypothesis of the study was evaluated using two learning groups: *Whole then Sections (WtS)*, and *Integrated 2D3D (2D3D)*. The participants in the two learning groups received the same treatment except for the program used for learning sectional anatomy.

After obtaining informed consent from the participants, the psychometric tests were conducted. Participants were assigned to a learning condition, and they were given instruction videos to watch in a quiet room with headphones. After watching the video, the participants used the instructional programs to demonstrate their ability to use the programs to an experimenter.

All participants learned whole anatomy first. The participants learned whole anatomy from the three different views (coronal, sagittal, and axial) in alternating trials. The order of presentation of the views was counterbalanced over participants. Participants continued cycling through trials until they reached 90% accuracy on tests in three successive trials.

Once the participants finished learning whole anatomy, they were given a sectional anatomy pre-test (Sections Test). This test was given to the participants before they began learning trials with sectional anatomy, to test their initial ability to transfer knowledge of whole anatomy to identifying sectional anatomy.

After taking the Sections Test, participants started learning sectional anatomy using one of the two learning programs (Sections Only or 2D3D), based on their learning

31

group assignment. Participants continued learning until they reached a high performance criterion; 90% accuracy on three successive trials. They were given tests of generalization of knowledge immediately after they completed learning sectional anatomy. The three generalization tests, Global Cues, Submit Structure, and Submit Name, were presented in a counterbalanced order across participants. After the generalization tests, the participants were scheduled for their last session in the study 4-8 weeks from the day they finished learning neuroanatomy. On their last visit to the laboratory after the retention interval, participants were tested on the long-term retention test of sectional anatomy and the fourth generalization test (Numeric Cues), in that order. Participants received no feedback.

Balancing participants. Spatial ability is known to be an important predictor of learning in the spatial domain (Pani et al., 2005; Hegarty, Montello, Richardson, Ishikawa, & Lovelace, 2006; Cohen & Hegarty, 2007; Keehner et al, 2008; Chariker et al., 2011). Differences between groups on spatial ability can confound the interpretation of results. Therefore, an effort was made to balance the two groups on spatial ability by using DAT score when assigning participants to a learning group. Forty-eight participants were included in the initial analyses.

The participants also were tested on their ability to identify sectional anatomy, using the Sections Test, after they finished learning whole anatomy. Even though the participants in the two groups had comparable DAT scores, and they had received the same treatment in the experiment so far, it did not appear that the groups were matched on their Sections Test performance ($t = 1.43$, $p = 0.16$; $WtS = 62.9$, $SD = 11.34$; $2D3D = 58.16$, $SD = 11.73$; see Appendix A for group means and statistical values for DAT,

Designs test, and Sections Test scores). The Sections Test measures the ability to infer

sectional anatomy based on knowledge of 3D shape and organization. Thus, it is a type of

measure of spatial ability. If the Sections Test is effectively measuring differences in the

spatial ability of participants to transform 3D knowledge to 2D representation,

differences on the Sections Test performance between the two learning groups could

confound the interpretation of results. Another possibility is that this difference in

performance is related to mastery motivation of the participants, which is known to be

responsible for large variations in performance in spatial learning (Pani et al, 2005).

Even though the difference in Sections test performance was not significant at the

typical level of alpha, the two groups cannot be considered equivalent. Frick (1995)

pointed out that most psychological studies treat the absence of a statistically significant

difference (p level greater than 0.05) between two groups as an indicator of equivalence

(null hypothesis being true). However, two groups should be considered clearly matched

on a variable only when the p level is above 0.5 (Frick, 1995; Mervis & Klein-Tasman,

2004; Mervis & John, 2008).

Therefore, the two learning groups were matched on their Sections Test and DAT

scores by applying a matching algorithm. Participants who matched poorly were

excluded from statistical analyses. In particular, variation in DAT and Sections Test

scores were treated as dimensions in a Euclidean space (both varying from 1 to 100). The

standard distance formula (from the Pythagorean theorem) was used to calculate the

distance between participants in the space. Thus, for each *Whole then Sections*

participant, 24 values were calculated corresponding to the distance from each *Integrated*

2D3D participant. A low value indicated a good match, while a high value indicated a

poor match. For example, the best matched pair had corresponding DAT scores of 55 and

Sections Test scores of 63 and 64. In the use of the algorithm, the pair with the best

match was removed from consideration and the pair with the next lowest value was

identified. This continued until the worst matching pair of participants was identified.

These two participants were then removed from further statistical analyses. The

difference between the *Whole then Sections* and *Integrated 2D3D* groups were then

checked on DAT and Sections Test scores. This process was continued until the *p* level

for both subject variables were greater than 0.5. Twenty participants in each condition

were selected based on this algorithm. After the matching procedure, the two groups were

equivalent on their scores on the DAT, Designs Test, and Sections Test scores ($p > 0.5$;

Table 1). It should be noted that the statistical significance of critical comparisons did not

change by matching participants in this manner. However, after matching participants in

the two groups, we are more confident that the two groups were equivalent.

Table 1

The mean and standard deviation of the mean for DAT, Designs Test, and Sections Test for WtS and 2D3D after matching the groups.

After Matching (*N* = 40)	*WtS*	*2D3D*	**Statistical Values**
DAT	*M* = 75.2	*M* = 73.65	*t* = 0.950
	SD = 21.16	*SD* = 20.74	*p* = 0.816
Designs Test	*M* = 11.10	*M* = 11.25	*t* = -0.164
	SD = 3.06	*SD* = 2.73	*p* = 0.871
Sections Test	*M* = 62.09	*M* = 61.04	*t* = 0.352
	SD = 9.92	*SD* = 8.81	*p* =0.727

Scoring. Three of the generalization tests, Global Cues, Numeric Cues, and Submit Structure, required manual scoring of the locations of mouse clicks on images. The scoring was done with the help of biomedical images that had predefined boundaries for each structure. The individual who scored the responses was blind to the participant's experimental group.

Item difficulty analysis. A procedure for the analysis of item difficulty was developed in a previous work to identify particularly difficult test items in each test (Chariker et al., 2012). The item difficulty analysis was conducted using binomial logistic regression following these procedures:

1. The most basic measure of item difficulty was the proportion of the sample in the control group (*Whole then Sections*) that identified an item correctly.

2. Binomial logistic regression was used to determine the mean proportion correct and the upper and lower bounds of the mean at the 95% confidence interval for each test.

3. Test items above the lower bound of the mean proportion correct were categorized as typical, and items below the lower bound of the mean were categorized as difficult.

4. Mean performance of each participant on these typical and difficult test items was calculated for each test for both the learning groups.

RESULTS

The analyses of data involved correlation, multilevel binomial modeling, and

mixed randomized ANOVA. The binomial models were used to determine difficult and

typical test items for analysis of the data from the long-term retention and generalization

tests.

Learning Whole and Sectional Anatomy

Learning whole anatomy. Participants learned whole anatomy in relatively few

trials ($M = 4.98$ trials, $SD = 1.48$). As shown in Figure 12, the mean percent correct on

the first learning trial was 52.8% ($SD = 16.24$) and reached 95.7% ($SD = 7.8$) by the third

trial. As expected, there was no significant difference in the number of trials to learn

whole anatomy between the two groups, $t(38) = 0.963$, $p = 0.342$ (*WtS*: $M = 5.2$, $SD =$

1.88; *2D3D*: $M = 4.75$, $SD = 0.91$).

Sections test. After completing learning of whole anatomy, participants were

tested on their ability to identify sectional anatomy prior to instruction in sectional

anatomy. As discussed earlier, this test was also used to match the participants in the two

learning groups. The participants were quite good at transferring their knowledge of

whole anatomy to identifying sectional anatomy ($M = 60.5\%$, $SD = 11.66$).

Learning sectional anatomy. Mean percent correct across all learning trials for

the two sectional anatomy learning conditions is presented in Figure 12. The *Whole then*

Sections group performed at 79.0% correct (*SD* = 13.31) on their first learning trial, while the *Integrated 2D3D* group started at 73.7% correct (*SD* = 12.74). This difference in performance was not statistically significant, $t(38) = 1.30$, $p = 0.203$.

To further analyze if the learning groups differ in their performance across learning trials, a 5x 2 mixed randomized ANOVA was conducted with performance on trial 1 to trial 5 as a repeated measure variable, and learning condition as a between group variable. The analysis included only the first five trials because a large number of participants completed learning in five trials (*n* in trial 5: *WtS* = 17; *2D3D* = 18). Even in the first four trials, several participants finished learning. Data for these participants was extrapolated to five trials. The main effect of learning condition was not significant, $F(1, 38) = 2.58$, $p = 0.116$, $n_p^2 = 0.064$. The interaction of learning condition with performance on sectional anatomy trials also was not significant, $F(4, 152) = 206$, $p = 0.303$, $n_p^2 = 0.031$.
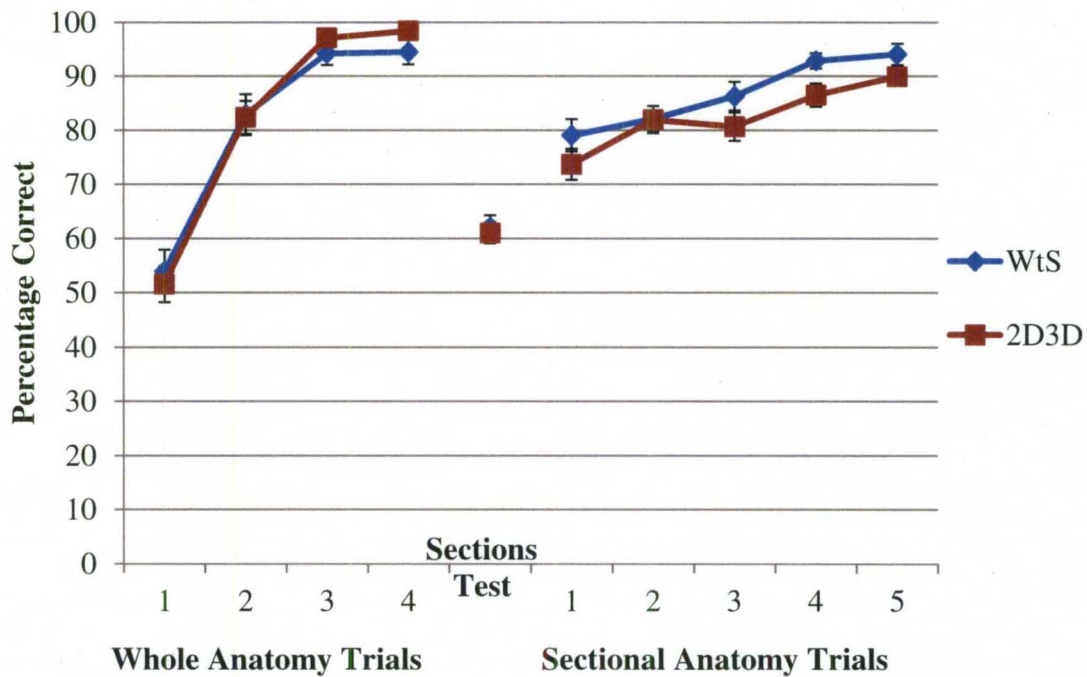
Figure 12. Percent correct performance during the learning stage for the two learning groups on the whole anatomy trials, the Sections Test, and the sectional anatomy learning trials. Error bars are standard error of the mean. Note: Some participants finished learning in three trials. The missing data was handled by extrapolating their performance on subsequent trials.

Participants were able to learn sectional anatomy to the criterion level in a few trials ($M = 8.57$, $SD = 4.37$). The mean number of sectional anatomy trials for the *Whole then Sections* condition was 7.8 ($SD = 4.83$) and the mean for the *Integrated 2D3D* condition was 9.35 ($SD = 3.82$). There was not a significant difference between the two learning groups in the number of trials to learn sectional anatomy, $t(38) = -1.13$, $p = 0.267$.

The order of presentation of the anatomical views was counterbalanced in this study, which permits us to explore which view was more challenging during sectional

anatomy learning. A one-way repeated measure ANOVA was conducted on the first three learning trials. There was a significant main effect of view, $F = 6.055$, $p = 0.004$, $n_p{}^2 = 13.4$. The performance on the coronal view was significantly better than performance on the sagittal and axial views (Coronal: $M = 84.06$, $SD = 10.09$; Sagittal: $M = 78.63$, $SD = 12.05$; Axial: $M = 79.19$, $SD = 14.01$; pair-wise comparisons using Bonferroni correction, Coronal vs. Sagittal: $MD = 5.43$, $p = 0.007$; Coronal vs. Axial: $MD = 4.87$, $p = 0.027$). The Sagittal and Axial views were not significantly different from each other.

Long-term Retention

The participants were tested on sectional anatomy after a retention interval of 4-8 weeks. The mean performance of the *Whole then Sections* group across the three anatomical views was 83.4%. The *Integrated 2D3D* group performed slightly better than the *Whole then Sections* group at 87.3%. However, this difference was not significant, $t(38) = -1.60$, $p = 0.119$.

An item difficulty analysis was conducted for the data from the test of long-term retention. This analysis led to categorization of 67 test items as typical (78.8% of the test items) and 18 as difficult (21% of the test items). The mean percent correct was calculated for each participant for the typical and the difficult items. The mean performance on typical and difficult test items for the two learning conditions are shown in Figure 13. A 2 x 2 mixed randomized ANOVA was conducted using item difficulty as a repeated measure variable and learning condition as a between group variable. There were significant main effects of item difficulty, $F(1, 38) = 206$, $p < 0.001$, $n_p{}^2 = 0.844$, and learning condition, $F(1, 38) = 6.171$, $p = 0.018$, $n_p{}^2 = 0.140$. The interaction between

item difficulty and condition also was significant, $F(1, 38) = 11.943$, $p = 0.001$, $n_p^2 = .239$. The pair-wise comparisons show a significant effect of condition for difficult test items, $F(1, 38) = 8.929$, $p = 0.005$, $n_p^2 = 0.190$. The *Integrated 2D3D* group was 17 percent better than the *Whole then Sections* group in remembering difficult test items. The performance on typical test items was not significantly different for the two learning groups (*Whole then Sections*: $M = 92.7$, $SD = 5.95$; *2D3D*: $M = 92.9$; $SD = 5.95$).
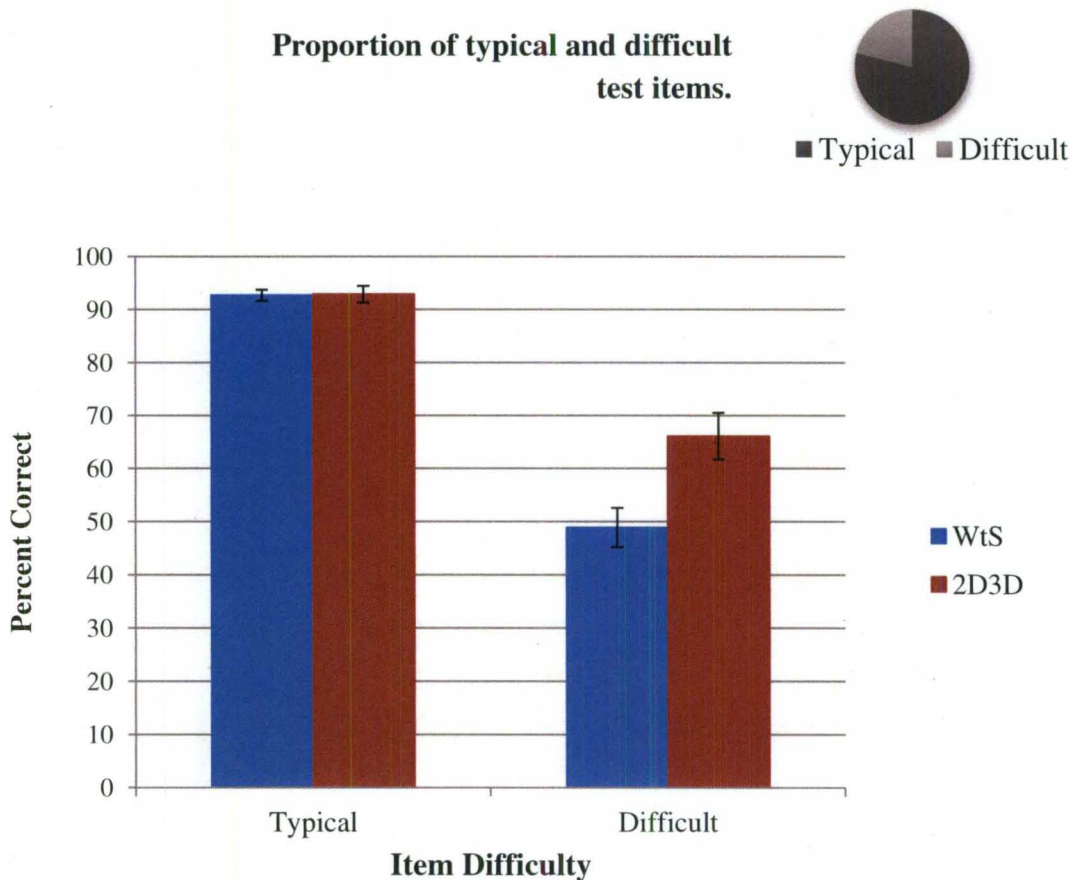
Figure 13. Mean percent correct on the test of long-term retention of sectional anatomy for the two learning groups for typical and difficult test items. Error bars show standard error of the mean.

41

A 2x2x2 mixed randomized ANOVA was conducted to analyze the role of spatial ability in learning from graphically integrated representation. Item difficulty, learning condition, and DAT were used as independent variables. The participants were divided by spatial ability into two groups using a median split: a higher spatial group ($M = 90.6$) and a lower spatial group ($M = 58.2$). The mean percentile for the lower spatial group was quite high (58 percentile), therefore this group should be considered average on spatial ability. There was no interaction of spatial ability with learning condition. Participants with both higher and lower spatial ability had similar performance on typical test items (approximately 93% for all four groups). Both the higher and lower spatial ability groups in *Integrated 2D3D* condition retained difficult test items at a higher level than higher and lower spatial groups in *Whole then Sections* condition (*WtS*: Higher - 53.9%; Lower - 43.9%; *2D3D*: Higher - 68.3%; Lower - 63.9%).

Generalization Tests: Interpretation of Biomedical Images

Three tests of generalization to interpreting new biomedical images were presented immediately after the learning trials. The order of presentation of these tests was counterbalanced. The fourth generalization test was presented after the long-term retention interval.

Effect of test type (cues). Another question of interest was how effective are the cues used in the tests of generalization. Therefore, the performance of the three generalization tests was compared collapsed across the learning groups. The mean percent correct on the Submit Name test was 70%, the mean percent correct on the Submit Structure test was 64%, and the mean percent correct on the Global Cues test was

52% (Figure 14). A repeated measure ANOVA was conducted on the three generalization tests (collapsed across order of presentation). There was a significant main effect of test-type, $F(2, 78) = 52.589$, $p < 0.001$, $n_p^2 = 0.574$. Performance on the Submit Name test was significantly better than performance on both the Global Cues, $t(39) = 10.92$, $p = 0.00$, $d = 1.96$, and the Submit Structure test, $t(39) = 3.15$, $p = 0.003$, $d = 0.59$. Performance on the Submit Structure test was significantly better than performance on the Global Cues test, $t(39) = 6.89$, $p = 0.00$, $d = 1.11$.
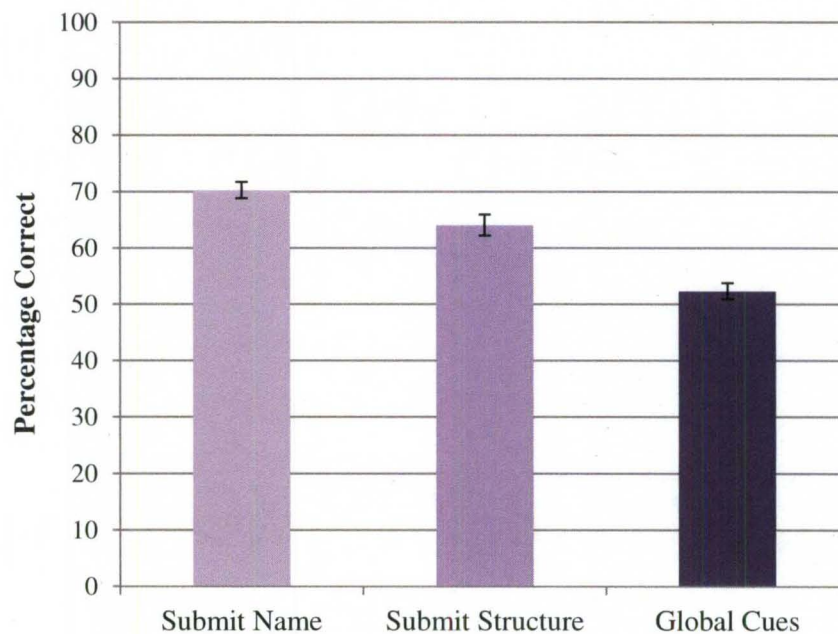
Figure 14. Mean percentage correct for the three generalization tests. Error bars are standard error of the mean.

Effect of learning condition. The item difficulty analysis was also conducted on the generalization tests. The performance on typical and difficult test items for the generalization tests are presented in Figure 15. A 2x2 mixed randomized ANOVA was conducted on each of the four generalization tests. The result of analyses is summarized in Table 2.
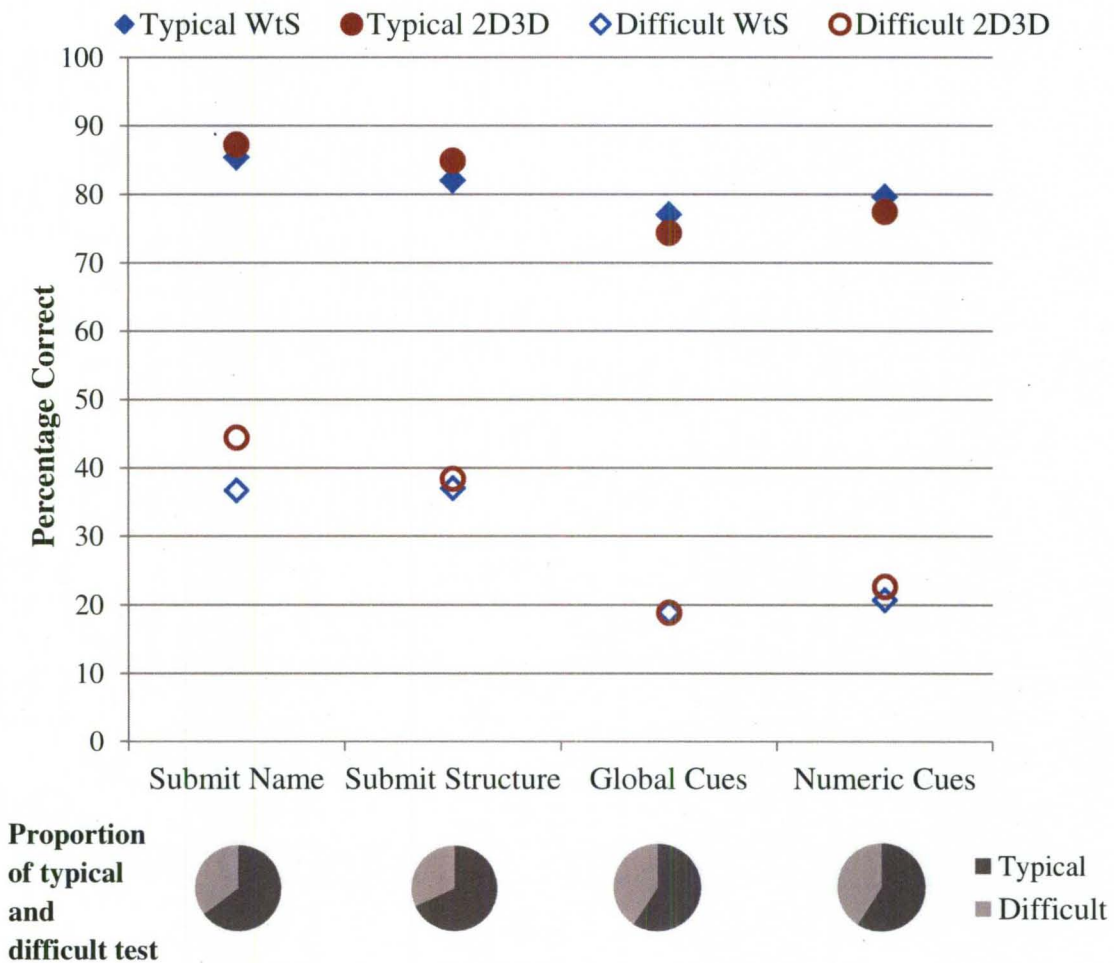
Figure 15. Performance on the generalization tests: Mean percent correct of the two learning groups is presented for typical and difficult test items for the three generalization tests: Submit Name, Submit Structure, Global Cues, and Numeric Cues.

In general, there were no effects of learning group or interaction of group with item difficulty. In particular, in the Submit Name test the item difficulty analysis led to categorization of 60 (35.3%) test items as difficult and 110 (64.7%) as typical test items. There was a main effect of item difficulty, $F(1,38) = 429.335$, $p < 0.001$, $n_p^2 = 0.919$. The mean of performance on typical test items was 86.8% ($SD = 6.3$) and performance on difficult test item was at 41.08% ($SD = 16.58$). There was no main effect for learning

condition. The interaction of learning condition with item difficulty was also not significant.

In the Submit Structure test, the item difficulty analysis led to categorization of 54 (31.2%) test items as difficult and 116 (68.2%) as typical test items. There was a main effect of item difficulty, $F(1,38) = 650.388$, $p < 0.001$, $n_p^2 = 0.945$. The mean of performance on typical test items was 83.7% ($SD = 8.7$) and performance on difficult test item was at 37.9% ($SD = 16.45$). There was no main effect for learning condition. The interaction of learning condition with item difficulty was also not significant.

In the Global Cues test, the item difficulty analysis led to categorization of 74 (43.5%) test items as difficult and 107 (62.9%) as typical test items. There was a main effect of item difficulty, $F(1,38) = 3149.024$, $p < 0.001$, $n_p^2 = 0.988$. The mean of performance on typical test items was 76.1% ($SD = 10$) and performance on difficult test item was at 19.1% ($SD = 9.09$). There was no main effect for learning condition. The interaction of learning condition with item difficulty was also not significant.

In the Numeric Cues test, the item difficulty analysis led to categorization of 54 (31.2%) test items as difficult and 116 (68.2%) as typical test items. There was a main effect of item difficulty, $F(1,38) = 650.388$, $p < 0.001$, $n_p^2 = 0.945$. The mean of performance on typical test items was 78.6% ($SD = 7.38$) and performance on difficult test item was at 21.7% ($SD = 8.7$). There was no main effect for learning condition. The interaction of learning condition with item difficulty was also not significant.

45

Table 2.

Summary of mixed randomized ANOVA analyses for the tests of generalization.

	Submit Name
Item Difficulty	$F(1,38) = 429.335, p < 0.001, n_p^2 = 0.919$
Condition	$F(1,38) = 1.805, p = 0.187, n_p^2 = 0.045$
Item Difficulty x Condition	$F(1,38) = 2.097, p = 0.156, n_p^2 = 0.052$
	Submit Structure
Item Difficulty	$F(1,38) = 650.388, p < 0.001, n_p^2 = 0.945$
Condition	$F(1,38) = 0.330, p = 0.569, n_p^2 = 0.009$
Item Difficulty x Condition	$F(1,38) = 0.172, p = 0.680, n_p^2 = 0.005$
	Global Cues
Item Difficulty	$F(1,38) = 3149.024, p < 0.001, n_p^2 = 0.988$
Condition	$F(1,38) = 0.214, p = 0.646, n_p^2 = 0.006$
Item Difficulty x Condition	$F(1,38) = 1.656, p = 0.206, n_p^2 = 0.042$
	Numeric Cues
Item Difficulty	$F(1,37) = 2141.4, p < 0.001, n_p^2 = 0.983$
Condition	$F(1,37) = 0.001, p = 0.970, n_p^2 = 0.000$
Item Difficulty x Condition	$F(1,37) = 2.894, p = 0.097\ n_p^2 = 0.073$

Effect of test order. The three generalization tests presented just after completion of learning were presented in a counterbalanced order. This allows us to explore the secondary hypothesis about *preparation for future learning*. A one-way repeated measure analysis was conducted on the effect of test position. In particular, the score of each participant on each of the three generalization tests was categorized according to the ordinal position in which it was taken (first, second, or third). There was a significant effect of test order, $F(2,78) = 12.034, p < 0.001, n_p^2 = 0.236$. As shown in Figure 16,

performance of the participants improved on the generalization tests as they moved from the first test ($M = 56$, $SD = 12.77$) to the second ($M = 63.2$, $SD = 10.7$) and the third test ($M = 67.5$, $SD = 11.2$). Pair-wise comparisons of the three test positions was conducted with Bonferroni correction ($p = 0.016$). The performance on the second and third tests was significantly better than the performance on the first test (first vs. second: $t(39)= 3.51$, $p = 0.001$, $d = 0.62$; first vs. third: $t(39) = 4.50$, $p = 0.00$, $d = 0.96$).
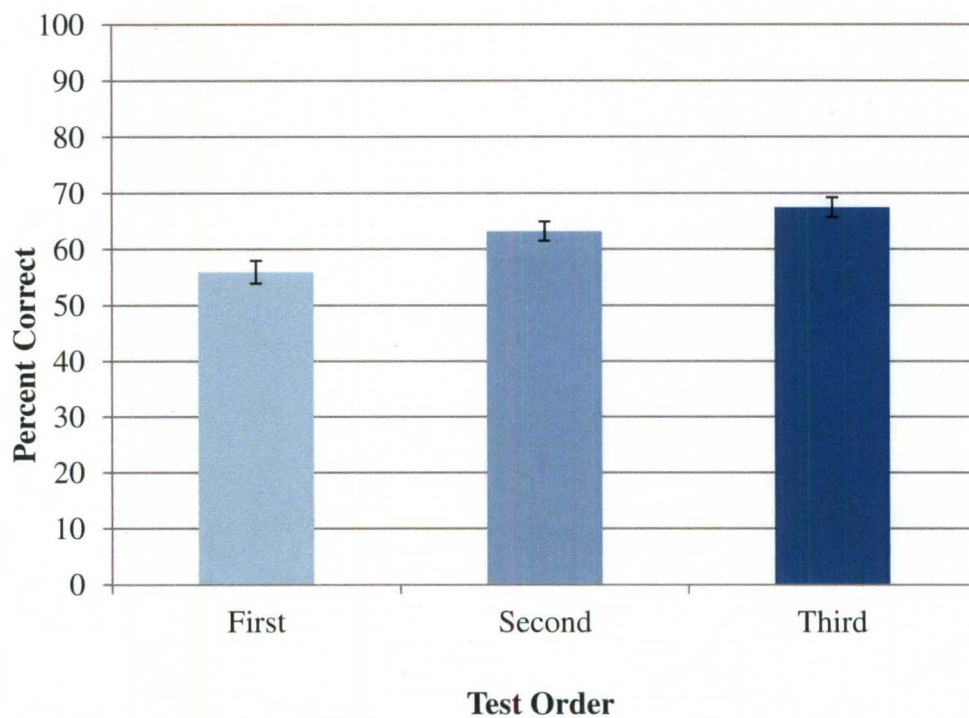
Figure 16. Mean percentage correct for the first, second, and third generalization tests. Error bars show the standard error of the mean.

Correlation of psychometric test scores

A correlation matrix was constructed to examine the relationships among the psychometric test scores and the outcome measures that did not depend on the experimental treatment (see Table 3). The outcome measures were the number of trials

required to complete whole anatomy learning and performance on the Sections Test. DAT scores were correlated with Spatial Memory scores ($r = 0.408$, $p = 0.009$) and Overall Memory scores ($r = 0.421$, $p = 0.007$). The Visual Content scores were not correlated with DAT ($r = 0.243$, $p = 0.131$). As expected, the Overall Memory scores were highly correlated with the Visual Content ($r = 0.853$, $p < 0.001$) and the Spatial Memory scores ($r = 0.669$, $p < 0.001$). Because Spatial Memory and Visual Content memory scores are sub-scores of Overall memory, and they are highly correlated with each other, these measures will only be discussed with outcome measures if and when the Overall scores are not significantly correlated but the Spatial Memory or Visual Content memory are.

Table 3.

Correlation Matrix (Pearson) relating psychometric test scores with number of whole and sectional anatomy trials and Sections Test.

	Overall	Visual_C	Spatial_M	Trials_WA	Sections_T
DAT	**.421	.243	**.408	***-.581	*.345
Overall		**.853	***.699	*-.333	**.515
Visual_C			*.381	-.230	**.452
Spatial_M				-.144	**.411

Acronyms used in the table: Overall for Overall visuospatial short term memory, Visual_C for Visual Content memory, Spatial_M for Spatial Memory, Trials_WA for number of whole anatomy trials, and Sections_T for Sections Test performance. Asterisk indicate statistical significance: * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$.

48

Learning whole anatomy. Spatial ability and the Overall Memory score for visuospatial memory were correlated with the number of trials to learn whole anatomy (DAT: $r = -0.581$, $p < 0.001$; Overall: $r = -0.333$, $p < 0.05$; see Table 3). The DAT and Overall Memory scores are significantly correlated with each other. In order to estimate their independent correlation with the number of whole anatomy trials, partial correlation was calculated for both the psychometric test scores. Partial correlation of DAT scores with the number of trials to learn whole anatomy after controlling for Overall Memory scores was significant ($r = -0.516$ $p < 0.01$). Partial correlation of Overall Memory scores with the number of trials to learn whole anatomy after controlling for DAT was not significant ($r = -0.119$ $p = 0.469$). In other words, DAT shares a substantial amount of variance with Overall visuospatial memory in accounting for the number of trials to learn whole anatomy. In addition, DAT contributes uniquely to accounting for the number of trials.

Sections test. The Sections Test scores were correlated with all of the psychometric test measures (see Table 3).The performance on the Sections Test correlated most highly with Overall Memory ($r = 0.515$, $p = 0.001$). Due to shared variance between Overall visuospatial short-term memory and DAT, partial correlations were estimated for Sections Test performance after controlling for each of these variables. Partial correlation of Overall Memory scores with Sections Test scores after controlling for DAT was significant ($r = 0.434$ $p < 0.01$). Partial correlation of DAT with Sections Test scores after controlling for Overall Memory scores was not significant ($r = 0.165$ $p = 0.315$). Thus, for the Sections Test, there was substantial shared variance

49

among the statistical predictors, with Overall visuospatial short-term memory providing a unique ability to account for Sections Test performance.

Learning sectional anatomy. The correlations of psychometric test scores with the number of trials to learn sectional anatomy were computed after removing two outliers from the data. Because the two groups learned sectional anatomy using different programs, the correlations were computed separately for the two groups. For the *Whole the Sections* group, the number of trials to learn sectional anatomy was correlated with DAT ($r = -0.526, p < 0.05$). Overall Memory ($r = -0.356, p = .135$), Content Memory ($r = -0.291, p = 0.227$), and Spatial Memory scores ($r = -0.443, p = 0.057$) were not correlated with the number of trials to learn.

For the *Integrated 2D3D* group, the number of trials to learn sectional anatomy was correlated with the Overall Memory scores ($r = -0.532, p < 0.05$) and the Spatial Memory scores ($r = -0.527, p < 0.05$). The DAT ($r = -0.404, p = 0.086$) and Content Memory scores ($r = -0.319, p = 0.184$) were not correlated.

Long-term retention. The correlations of psychometric test scores with performance on the test of long-term retention were measured separately for each learning group. For both groups, long-term retention was not correlated with any of the psychometric measures (*WtS*: DAT: $r = 0.332, p = 0.152$; Overall: $r = 0.203, p = 0.392$; Visual Content: $r = 0.097, p = 0.684$; Spatial Memory: $r = 0.204, p = 0.388$; *2D3D*: DAT: $r = -0.023, p = 0.922$; Overall: $r = 0.173, p = 0.466$; Visual Content: $r = 0.168, p = 0.478$; Spatial Memory: $r = 0.226, p = 0.339$).

The correlation of performance in long-term retention broken down by item difficulty with the psychometric scores was computed for both the learning groups. There was a significant correlation between performance on the difficult test items and the DAT scores for the *Whole then Sections* group only (WtS: $r = .445$, $p < 0.05$; *2D3D*: $r = .099$, $p = 0.676$). The complete correlation matrix of psychometric tests with item difficulty for the two learning groups is presented in Appendix B.

Tests of generalization of knowledge. The outcome measures of generalization tests were correlated with psychometric test scores. There were no significant correlations of interest in these correlations. The complete correlation matrix of psychometric tests with item difficulty for all four generalization tests is presented in Appendix C.

DISCUSSION

The computer-based learning programs used in this study were very good tools for learning neuroanatomy. Participants in both the learning groups learned whole and sectional anatomy in a few trials (approximately 14 trials). Performance on the Sections Test suggests that participants were able to transfer substantial amounts of whole anatomy knowledge to inferring sectional anatomy. Support for the primary hypothesis that learning from graphically integrated 2D and 3D representation would be a more effective way to learn sectional anatomy comes from the test of long-term retention. Learning from a graphically integrated model of 2D and 3D representation helped the participants in retaining the difficult sectional information better than learning from a sequential presentation. This finding shows that graphically integrated representations have a unique advantage of illustrating the 3D to 2D transformation and apparently helped the participants to build a more accurate mental model of neuroanatomy.

It is important to emphasize that this difference in performance was found only for the difficult test items. These test items appear to be challenging because they are confusable with other structures, or have an atypical appearance in some of the sections of the brain (Chariker et al., 2012). Thus, the *Whole then Sections* group was able to retain the knowledge of typical test items at a high level but struggled with a small set of difficult test items. The *Integrated 2D3D* group could see the transformation of 3D structures into 2D slices, which allowed them to have a deeper understanding of the spatial relationships among these structures.

52

This finding is consistent with the literature on spatial cognition, which suggests that imagining spatial transformations for even simple objects can be very challenging. If the objects are not in a canonical orientation, not symmetric, or are not aligned to a salient reference system, spatial imagination may be inaccurate (Hinton, 1979; Kozhevnikov, & Hegarty, 2001; Pani et al., 1996; Pani et al., 2005; Reed, 1974; Stevens & Coupe, 1978; Tversky, 1981). The results suggest that much spatial information about whole and sectional anatomy can be learned, transferred, and retained without explicit demonstration of spatial transformations. However, in some cases where attention to spatial detail is necessary, explicit presentation of the information to be integrated is more beneficial than sequential presentation.

Another important finding in long-term retention was the lack of interaction of spatial ability with the learning conditions. The literature on computer-based learning has expressed concern for the ability of low spatial learners to learn from complex visualizations. This study shows that the lower spatial group also benefited from an integrated presentation. While the higher spatial group in the *Integrated 2D3D* condition was 14% better than *Whole then Sections*, the lower spatial group in the *Integrated 2D3D* condition was 20% better than the lower spatial group in *Whole then Sections* condition. This finding is consistent with the studies which have used longitudinal learning sessions to train participants (Chariker, 2009; Keehner et al., 2006, Pani et al., 2005). Based on this finding, it can be argued that educational visualization tools can be beneficial to learners of varying spatial ability if the exposure to the material is not limited. With enough time, participants with lower spatial ability adapt to the complexity of graphically integrated visualization and benefit from it.

Support for the secondary hypothesis about *preparation for future learning* was also found. The analysis of test order for the tests of generalization showed that the participants were able to generalize their knowledge to a high level and continued learning with the help of cues from the first generalization test to the second and third. The MRI and Visible Human images used in these tests were very different from the neuroanatomy model used for learning. Evidence of good generalization and continued learning with these new representations of neuroanatomy suggests that the participants were proficient in their neuroanatomy knowledge. This finding is encouraging because in real world learning situations preparedness for future learning is considered important (Bransford & Schwartz, 1999).

Another interesting finding of the study was the effect of different cues in the tests of generalization of knowledge. It was expected that the more specific cues (such as arrows pointing to a structure, or the name of a structure that is present in the image) would be more helpful than global orienting cues. As expected, performance on the Submit Name and the Submit Structure tests was better than the performance on the Global Cues test. The cues in the Submit Name test were most helpful. This is probably because the arrows in the Submit Name test images were particularly helpful in eliminating from consideration the anatomical structures that the learner had not seen before.

An interesting outcome of the data analyses was the correlation between the two psychometric test scores ($r = .421$). These tests were measuring spatial ability using different tasks. The Differential Aptitude Test involves identification of the correct 3D object that can be made from folding the 2D pattern provided. The task suggests that it

would require thinking about manipulating an object in 2D and 3D space. The Designs

Test explores memory for patterns and spatial locations in a 2D grid. The task in the

Designs test involves thinking about spatial location in 2D space and visuospatial

memory for patterns. It is possible that the correlation between the two tests of spatial

memory is due to a common aspect of spatial ability that they measure: thinking in 2D

space.

There were some interesting correlations of performance outcome with the

psychometric tests. There was a significant correlation of DAT with the number of trials

to learn whole anatomy after controlling for the effect of Overall visuospatial memory

scores. It is possible that this unique relation of DAT with the number of trials to learn

whole anatomy occurs because learning whole anatomy requires participants to think in

3D space. The significant correlation of the Sections Test with the Overall Memory

scores after controlling for the effect of DAT suggests that inferring the 2D

representation with the help of whole anatomy knowledge largely involved thinking in

2D space. It will be important to see whether this result replicates in future work.

The two learning groups were not significantly different on their test performance

on the first five trials of sectional anatomy learning. Similarly, there were no differences

in the number of sectional anatomy trials to complete learning. However, the *Integrated

2D3D* group was 4.3% lower than the *Whole then Sections* group in their mean

percentage correct on the first five trials of sectional anatomy learning. This difference

approached statistical significance ($p = 0.116$). Similarly, the *Integrated 2D3D* group

took 1.6 more trials to learn sectional anatomy. At face value, it seemed possible that the

17% increase in performance of the *Integrated 2D3D* group for difficult test items on the

long-term retention test was due to the 1.6 additional trials taken by this group. To understand the role of number of trials in increasing retention, a comparison of the *Whole then Sections* group was done across the three studies in this project. The *Whole then Sections* group in the previous two studies learned using a more stringent criterion (90% or more performance on trials in six tests). Therefore, the number of trials to learn neuroanatomy for these groups was higher than the current study. In Figure 17, the performance of the three *Whole then Sections* groups on difficult test items in long-term retention test is plotted against number of trials taken by the groups to learn neuroanatomy (both whole and sectional anatomy). As shown in Figure 17, with the increase in number of trials, the performance of *Whole then Sections* increased across the three studies. However, only the *Whole then Sections* group in Chariker et al. (2012) had comparable performance to the *Integrated 2D3D* group. This group required approximately 12 more trials than the *Integrated 2D3D* group (nearly double) to learn neuroanatomy. This comparative look at the data suggests that the better performance of the *Integrated 2D3D* group on difficult test items cannot be explained by the two additional trials taken by the group to learn sectional neuroanatomy. Instead, the benefit appears to be due to the explicit graphical integration of 3D and 2D representation.
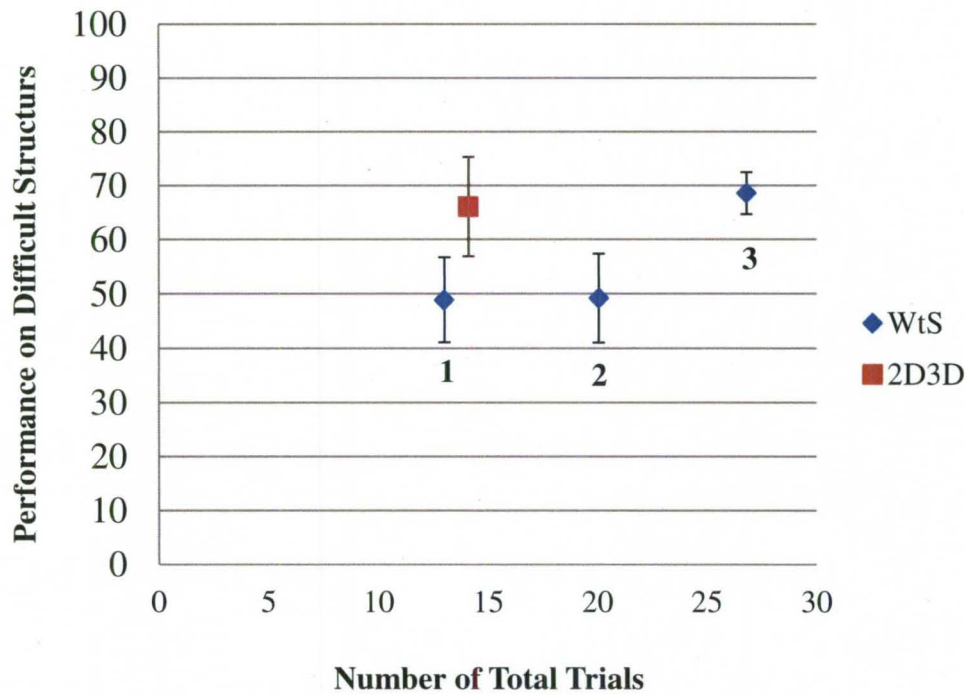
Figure 17. The performance of the learning groups on difficult test items in long-term retention test as a function of number of trials to learn neuroanatomy. The 95% Confidence Interval of the mean is plotted for performance on difficult test items. 1: *Whole then Sections* group in current study, 2: *Whole then Sections* group in Pani et al, 2012; 3: *Whole then Sections* in Chariker et al., 2011.

Limitations

The results of this study are promising for suggesting directions for developing computer-based instruction for learning neuroanatomy. The lack of interaction of spatial ability with learning condition is very encouraging. However, the sample mean percentile for spatial ability was 74.4, a rather high value. The trend in results suggests that the *Integrated 2D3D* should be beneficial for even very low percentile group, but this claim needs to be experimentally verified.

Another limitation of this study was that the participants showed good generalization for the typical test items, but struggled with the difficult test items. It appears that participants were unable to use their knowledge to orient themselves and reason about the possible structures present in the image. For example, the size of the cerebellum and the presence or absence of ventricles in an image should have provided them with information about where the slice was taken from the brain. However, the participants found this task very challenging. One possible reason for a lack of successful reasoning and problem solving in tests of generalization is that participants never had to rely solely on reasoning to identify structures during the learning stage of sectional anatomy. The sectional anatomy test slices were always presented in the same order (for instance, if the fourth, seventh, and tenth slices that participants sees during learning were selected for testing, they were always presented in ascending order). It is possible that randomizing the order of test slices would have made the participants think about the location of test slices in the model and focus more on using familiar structures as landmarks to reason about unknown structures. Another possibility is that the images used in tests of generalization are too complex to learn without explicit instruction. Exposure to these images during learning may be necessary for high levels of recognition.

CONCLUSIONS AND FUTURE PROSPECTS

The results of this study suggest that there are certain spatial relationships and transformations which require explicit demonstration. The graphically integrated visualization tools used in this study provided a unique advantage of presenting the spatial information explicitly. In domains such as medicine, the opportunity for learners to see the 3D and 2D mapping in representation of anatomy is very limited due to constraints on resources (e.g., dissection labs). Similar challenges may be encountered in the domain of chemistry for learning molecular organization, in physics for understanding organization at the atomic level or at the level of the universe, in biology for understanding the microscopic views of organisms, and in geography for understanding the 2D and 3D layout. In these domains, a real-world opportunity to see the spatial relationships and transformations is not even possible. A computer-based graphical visualization tool can be very beneficial in these cases.

The learning programs used in this study show the capability to make naïve learners proficient in neuroanatomy where they can continue to learn with some guidance. A good learning environment (real or virtual) should prepare the learner to adapt and tackle new problems in the domain effectively. Computer-based learning programs are usually evaluated on immediate learning outcomes. Based on findings of this study it can be suggested that computer-based learning programs need to adopt a more rigorous set of tests for evaluating their effectiveness beyond immediate learning

outcomes before being introduced in a curriculum to ensure it is providing some unique benefit.

An important point to remember when recommending integrated 2D and 3D graphics for visualization is that in this study this tool was beneficial for a small set of spatially challenging representations. It is possible that a similar benefit could be achieved by exposing participants to integrated graphics only for the difficult test items. For instance, a learning program can be created which uses just sectional anatomy in the study stage and the integrated 2D3D for feedback to allow participants to see where they made errors and why (3D to 2D transformations). Such a program may be more efficient than the existing approach while maintaining the benefits. Therefore, it would be important to evaluate the visualization tools to establish which approach to implementing graphical visualization would lead to good retention and generalization while retaining efficiency in learning.

REFERENCES

Ackerman, M. J. (1995). Accessing the Visible Human Project. D-Lib Magazine.

Available: http://www.dlib.org/dlib/october95/10ackerman.html

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with

multiple representations. *Learning and Instruction, 16*(3), 183-198.

Bennett, G.K., Seashore, H. G., & Wesman, A. G. (1989). *Differential Aptitude Tests for*

Personnel and Career Assessment: Space Relations. San Antonio, TX: The

Psychological Corporation, Harcourt Brace Jovanovich.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human

beings. In J. Metcalfe and A. Shimamura (Eds.), *Metacognition: Knowing about*

knowing(pp.185-205). Cambridge, MA: MIT Press.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with

multiple implications. *Review of Research in Education, 24 1999, 24,* 61-100. doi:

10.2307/1167267

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn.*

Washington, D.C.: National Academy Press.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval

schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior, 8*(3), 323-343. doi: 10.1016/s0022-5371(69)80124-6

Chariker, J. H. (2009). *Computer-based learning in neuroanatomy: A longitudinal study of learning, transfer, and retention.* (Ph.D. Dissertation), University of Louisville, Louisville. Retrieved from ProQuest Dissertations and Thesis database. (Accession number: AAT 3370062)

Chariker, J. H., Naaz, F., & Pani, J. R. (2011). Computer-Based Learning of Neuroanatomy: A Longitudinal Study of Learning, Transfer, and Retention. *Journal of Educational Psychology, 103*(1), 19-31. doi: 10.1037/a0021680

Chariker, J. H., Naaz, F., & Pani, J. R. (2012). Item difficulty in the evaluation of computer-based instruction: An example from neuroanatomy. *Anatomical Sciences Education, 5*(2), 63-75. doi: 10.1002/ase.1260

Codd, A. M., & Choudhury, B. (2011). Virtual Reality Anatomy: Is it Comparable with Traditional Methods in the Teaching of Human Forearm Musculoskeletal Anatomy? *Anatomical Sciences Education, 4*(3), 119-125.

Cohen, C. A., & Hegarty, M. (2007). Individual differences in use of external visualisations to perform an internal visualisation task. *Applied Cognitive Psychology, 21*(6), 701-711.

Cook, D. A., Levinson, A. J., & Garside, S. (2011). Method and reporting quality in health professions education research: a systematic review. *Medical Education,*

45(3), 227-238. doi: 10.1111/j.1365-2923.2010.03890.x

Cottam, W. W. (1999). Adequacy of medical school gross anatomy education as perceived by certain postgraduate residency programs and anatomy course directors. *Clinical Anatomy, 12*(1), 55-65. doi: 10.1002/(SICI)1098-2353(1999)12:1<55::AID-CA8>3.0.CO;2-O

Crowley, R. S., & Medvedeva, O. (2006). An intelligent tutoring system for visual classification problem solving. *Artificial Intelligence in Medicine, 36*(1), 85-117.

Dev, P. (1999). Imaging and visualization in medical education. *Ieee Computer Graphics and Applications, 19*(3), 21-31.

Dev, P., Hoffer, E. P., & Barnett, G. O. (2006). Computers in medical education. In E. H. Shortliffe & J. J. Cimino (Eds.), *Biomedical Informatics- Computer applications in health care and biomedicine* (Third ed., pp. 737-762). United States of America: Springer Science + Business Media, LLC.

Fitzgerald, J. E. F., White, M. J., Tang, S. W., Maxwell-Armstrong, C. A., & James, D. K. (2008). Are We Teaching Sufficient Anatomy at Medical School? The Opinions of Newly Qualified Doctors. *Clinical Anatomy, 21*(7), 718-724. doi: 10.1002/ca.20662

Frick, R. W. (1995). Accepting the null hypothesis. *Memory & Cognition, 23*(1), 132-138. doi: 10.3758/bf03210562

Garg, A., Norman, G. R., Spero, L., & Maheshwari, P. (1999). Do virtual computer

models hinder anatomy learning? *Academic Medicine, 74*(10), S87-S89.

doi: 10.1097/00001888-199910000-00049

Garg, A. X., Norman, G., & Sperotable, L. (2001). How medical students learn spatial

anatomy. *Lancet, 357*(9253), 363-364. doi: 10.1016/S0140-6736(00)03649-7

Garg, A. X., Norman, G. R., Eva, K. W., Spero, L., & Sharan, S. (2002). Is there any real

virtue of virtual reality?: The minor role of multiple orientations in learning

anatomy from computers. *Academic Medicine, 77*(10), S97-S99.

doi: 10.1097/00001888-200210001-00030

Gilbert, J.K.(Ed.) (2007). *Visualization in science education.* Dordecht, The Netherlands:

Springer.

Hariri, S., Rawn, C., Srivastava, S., Youngblood, P., & Ladd, A. (2004). Evaluation of a

surgical simulator for learning clinical anatomy. *Medical Education, 38*(8), 896-

902. doi: 10.1111/j.1365-2929.2004.01897.x

Hegarty, M. (2004). Dynamic visualizations and learning: getting to the difficult

questions. *Learning and Instruction, 14*(3), 343-351. doi:

10.1016/j.learninstruc.2004.06.007

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006).

Spatial abilities at different scales: Individual differences in aptitude-test

performance and spatial-layout learning. *Intelligence, 34*(2), 151-176. doi:

10.1016/j.intell.2005.09.005

Hinton, G. (1979). Some demonstrations of the effects of structural descriptions in mental

   imagery. *Cognitive Science, 3*(3), 231-250. doi: 10.1016/s0364-0213(79)80008-7

Hoffler, T. N. (2010). Spatial Ability: Its Influence on Learning with Visualizations-a

   Meta-Analytic Review. *Educational Psychology Review, 22*(3), 245-269. doi:

   10.1007/s10648-010-9126-7

Huk, T. (2006). Who benefits from learning with 3D models? the case of spatial ability.

   Journal of Computer Assisted Learning, 22(6), 392-404. doi: 10.1111/j.1365-

   2729.2006.00180.x

Issenberg, S. B., McGaghie, W. C., Petrusa, E. R., Gordon, D. L., & Scalese, R. J. (2005).

   Features and uses of high-fidelity medical simulations that lead to effective

   learning: a BEME systematic review. *Medical Teacher, 27*(1), 10-28. doi:

   10.1080/01421590500046924

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for

   learning. *Science, 319*(5865), 966-968. doi: 10.1126/science.1152408

Keedy, A. W., Durack, J. C., Sandhu, P., Chen, E. M., O'Sullivan, P. S., & Breiman, R. S.

   (2011). Comparison of Traditional Methods with 3D Computer Models in the

   Instruction of Hepatobiliary Anatomy. *Anatomical Sciences Education, 4*(2), 84-

   91. doi: 10.1002/ase.212

Keehner, M., Hegarty, M., Cohen, C., Khooshabeh, P., & Montello, D. R. (2008). Spatial

   Reasoning With External Visualizations: What Matters Is What You See, Not

Whether You Interact. *Cognitive Science, 32*(7), 1099-1132. doi:

10.1080/03640210801898177

Keehner, M., Lippa, Y., Montello, D. R., Tendick, F., & Hegarty, M. (2006). Learning a

spatial skill for surgery: How the contributions of abilities change with practice.

Applied Cognitive Psychology, 20(4), 487-503.

Khalil, M. K., Paas, F., Johnson, T. E., & Payer, A. F. (2005). Interactive and dynamic

visualizations in teaching and learning of anatomy: a cognitive load perspective.

Anatomical record. Part B, New anatomist, 286(1), 8-14. doi: 10.1002/ar.b.20077

Kikinis, R., Shenton, M. E., Iosifescu, D. V., McCarley, R. W., Saiviroonporn, P.,

Hokama, H. H., ... Jolesz, F. A. (1996). A digital brain atlas for surgical planning,

model driven segmentation and teaching. *IEEE Transactions on Visualization and*

Computer Graphics, 2, 232-241. doi:10.1109/2945.537306

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the

"Enemy of Induction"? *Psychological Science, 19*(6), 585-592. doi:

10.1111/j.1467-9280.2008.02127.x

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful Retrieval Attempts

Enhance Subsequent Learning. *Journal of Experimental Psychology-Learning*

Memory and Cognition, 35(4), 989-998. doi: 10.1037/a0015729

Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object manipulation

spatial ability and spatial orientation ability. *Memory & Cognition, 29*(5), 745-

756. doi: 10.3758/bf03200477

Lee, H. S., Linn, M. C., Varma, K., & Liu, O. L. (2010). How Do Technology-Enhanced

Inquiry Science Units Impact Classroom Learning? *Journal of Research in*

Science Teaching, 47(1), 71-90. doi: 10.1002/tea.20304

Levinson, A. J., Weaver, B., Garside, S., McGinn, H., & Norman, G. R. (2007). Virtual

reality and brain anatomy: a randomised trial of e-learning instructional designs.

Medical Education, 41(5), 495-501. doi: 10.1111/j.1365-2929.2006.02694.x

Lowe, R. (2004). Interrogation of a dynamic visualization during learning. *Learning and*

Instruction, 14(3), 257-274. doi: 10.1016/j.learninstruc.2004.06.003

Luursema, J. M., Verwey, W. B., Kommers, P. A. M., & Annema, J. H. (2008). The role

of stereopsis in virtual anatomical learning. *Interacting with Computers, 20*(4-5),

455-460. doi: 10.1016/j.intcom.2008.04.003

Luursema, J. M., Verwey, W. B., Kommers, P. A. M., Geelkerken, R. H., & Vos, H. J.

(2006). Optimizing conditions for computer-assisted anatomical learning.

Interacting with Computers, 18(5), 1123-1138. doi: 10.1016/j.intcom.2006.01.005

Melton, A. W. (1970). Situation with respect to spacing of repetitions and memory.

Journal of Verbal Learning and Verbal Behavior, 9(5), 596-606. doi:

10.1016/s0022-5371(70)80107-4

Mervis, C. B., & John, A. E. (2008). Vocabulary abilities of children with Williams

syndrome: Strengths, weaknesses, and relation to visuospatial construction ability.

Journal of Speech Language and Hearing Research, 51(4), 967-982. doi:

10.1044/1092-4388(2008/071)

Mervis, C. B., & Klein-Tasman, B. P. (2004). Methodological issues in group-matching

designs: alpha levels for control variable comparisons and measurement

characteristics of control and target variables. *Journal of Autism and*

Developmental Disorders, 34(1), 7-17. doi:

10.1023/B:JADD.0000018069.69562.b8

Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year

review of empirical research (1999-2009). *Computers & Education, 56*(3), 769-

780. doi: 10.1016/j.compedu.2010.10.020

Miller, C. S., Lehman, J. F., & Koedinger, K. R. (1999). Goals and learning in

microworlds. *Cognitive Science, 23*(3), 305-336. doi:

10.1207/s15516709cog2303_2

Nguyen, N., Nelson, A. J., & Wilson, T. D. (2012). Computer visualizations: Factors that

influence spatial anatomy comprehension. *Anatomical Sciences Education, 5*(2),

98-108. doi: 10.1002/ase.1258

Nicholson, D. T., Chalk, C., Funnell, W. R. J., & Daniel, S. J. (2006). Can virtual reality

improve anatomy education? A randomised controlled study of a computer-

generated three-dimensional anatomical ear model. *Medical Education, 40*(11),

1081-1087. doi: 10.1111/j.1365-2929.2006.02611.x

Pani, J. R., Chariker, J. H., Dawson, T. E., & Johnson, N. (2005). Acquiring new spatial intuitions: Learning to reason about rotations. *Cognitive Psychology, 51*(4), 285-333. doi: 10.1016/j.cogpsych.2005.06.002

Pani, J.R., Chariker, J.H., & Naaz, F (2012). *Computer based learning: Interleaving whole and sectional representation of neuroanatomy.* Manuscript submitted for publication.

Pani, J. R., Jeffres, J. A., Shippey, G. T., & Schwartz, K. J. (1996). Imagining projective transformations: Aligned orientations in spatial organization. *Cognitive Psychology, 31*(2), 125-167. doi: 10.1006/cogp.1996.0015

Ratiu, P., Hillen, B., Glaser, J., & Jenkins, D. P. (2003). Visible Human 2.0: The next generation. In J.D.Westwood, H. M. Hoffman, G. T. Mogel, R. Phillips, R. A. Robb, & D. Stredney (Eds.), *Medicine Meets Virtual Reality 11 - NextMed: Health Horizen* (pp. 275-281). Amsterdam, The Netherlands: IOS Press.

Reed, S. K. (1974). Structural descriptions and limitations of visual images. *Memory & Cognition, 2*(2), 329-336. doi: 10.3758/bf03209004

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science, 35*(6), 481-498. doi: 10.1007/s11251-007-9015-8

Russell, J. W., & Kozma, R. B. (1994). 4MCHEM - Multimedia and mental models in chemistry. *Journal of Chemical Education, 71*(8), 669-670.

Silen, C., Wirell, S., Kvist, J., Nylander, E., & Smedby, O. (2008). Advanced 3D

visualization in student-centred medical education. *Medical Teacher, 30*(5), E115-E124. doi: 10.1080/01421590801932228

Stevens, A., & Coupe, P. (1978). Distortions in judged spatial relations. *Cognitive Psychology, 10*(4), 422-437. doi: 10.1016/0010-0285(78)90006-3

Taylor, K., & Rohrer, D. (2010). The Effects of Interleaved Practice. *Applied Cognitive Psychology, 24*(6), 837-848. doi: 10.1002/acp.1598

Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology, 13*(3), 407-433. doi: 10.1016/0010-0285(81)90016-5

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate? *International Journal of Human-Computer Studies, 57*(4), 247-262. doi: 10.1006/ijhc.1017

Waterston, S. W., & Stewart, I. J. (2005). Survey of clinicians' attitudes to the anatomical teaching and knowledge of medical students. *Clinical Anatomy, 18*(5), 380-384. doi: 10.1002/ca.20101

Wechsler D. (2009). *Wechsler Memory Scale* (fourth edition). San Antonio, Texas: Pearson Education, Inc.

APPENDIX A

The mean and standard deviation of the mean for DAT, Designs Test, and Sections Test for WtS and 2D3D before after matching the groups.

Before Matching ($N = 48$)	WtS	2D3D	Statistical Values
DAT	$M = 73.17$	$M = 74.75$	$t = -0.248$
	$SD = 23.55$	$SD = 20.53$	$p = 0.805$
Designs Test	$M = 11.13$	$M = 11.04$	$t = 0.108$
	$SD = 2.79$	$SD = 2.58$	$p = 0.915$
Sections Test	$M = 62.91$	$M = 58.16$	$t = 1.426$
	$SD = 11.34$	$SD = 11.73$	$p = 0.160$
After Matching ($N = 40$)			
DAT	$M = 75.2$	$M = 73.65$	$t = 0.950$
	$SD = 21.16$	$SD = 20.74$	$p = 0.816$
Designs Test	$M = 11.10$	$M = 11.25$	$t = -0.164$
	$SD = 3.06$	$SD = 2.73$	$p = 0.871$
Sections Test	$M = 62.09$	$M = 61.04$	$t = 0.352$
	$SD = 9.92$	$SD = 8.81$	$p = 0.727$

APPENDIX B

Correlation Matrix (Pearson) relating psychometric test scores with long-term retention test of sectional anatomy broken by item difficulty.

| | WtS | | 2D3D | |
	Typical	Difficult	Typical	Difficult
DAT	.125	*.445	-.112	.099
Overall	.289	.054	.151	.171
Visual_C	.267	-.109	.152	.157
Spatial_M	.129	.228	.188	.234

Overall for Overall visuospatial short term memory, Visual_C for Visual Content memory, and Spatial_M for Spatial Memory. Asterisk indicate statistical significance: * for $p < 0.05$.

APPENDIX C

Correlation Matrix (Pearson) relating psychometric test scores with tests of generalization of knowledge broken by item difficulty.

Typical test items	WtS				2D3D			
	SN	SS	GC	NC	SN	SS	GC	NC
DAT	.41	.35	-.25	.42	.27	-.02	.27	-.12
Overall	.19	.03	-.32	.21	.11	.14	*.44	.41
Visual_C	.29	-.20	-.40	-.02	.05	.07	.24	.36
Spatial_M	-.23	.04	-.41	.23	.17	.35	*.48	.40

Difficult test items	WtS				2D3D			
	SN	SS	GC	NC	SN	SS	GC	NC
DAT	.28	.32	.17	*.47	.23	.31	.25	.05
Overall	.15	.24	.20	.14	.21	*.53	*.45	**.60
Visual_C	.05	.04	.06	.06	.02	.40	.20	*.53
Spatial_M	.06	.02	-.05	.01	.22	*.50	*.46	.32

Acronyms used in the table: Overall for Overall visuospatial short term memory, Visual_C for Visual Content memory, Spatial_M for Spatial Memory, SN for Submit Name, SS for Submit Structure, GC for Global Cues, and NC for Numeric Cues. Asterisk indicate statistical significance: * for $p < 0.05$ and ** for $p < 0.01$.

CURRICULUM VITAE

Contact Information:

Department of Psychological and Brain Sciences
317 Life Sciences Building
University of Louisville, Louisville, KY 40292
Phone: 502- 435- 3263
Email: farah.naaz@louisville.edu

Education:

2006 - Present   Ph.D. Candidate, Experimental Psychology, University of
     Louisville.
     Dissertation Title: Learning from graphically integrated 2D and 3D
     representations improves retention of neuroanatomy.

2006   M.Sc. in Cognitive Science, Center for Behavioural and Cognitive
     Sciences, University of Allahabad.
     Thesis Title: Change detection and attention.

2004   B.Sc., University of Allahabad.

Fellowships:

Fall 2011   Doctoral Dissertation Completion Award, School of
     Interdisciplinary and Graduate Studies, University of Louisville

2006 - 2008   University Fellowship, School of Interdisciplinary and Graduate
     Studies, University of Louisville

2004 - 2006   Merit Scholarship, Center for Behavioural and Cognitive Sciences,
     University of Allahabad

Research Interests:

Visual Cognition; Learning and Memory; Computer-Based Learning

Research Experience:

Fall 2007 - Present   Research Assistant, Visual Cognition Lab*

* Supported by a grant from the National Library of Medicine, National Institutes of Health (1 R01 LM008323; Histological Reasoning: Visual Cognition in Microanatomy; PI: J.R. Pani) in Fall 2008 & Fall 2009.

Teaching Experience:

Spring 2012   Instructor, Sensation and Perception, University of Louisville.

Spring 2011   Co-Instructor, Sensation and Perception, University of Louisville.

Fall 2010   Teaching Assistant, Experimental Psychology, University of
                     Louisville. Lab sections involved instructing students on aspects of
                     research design, writing research reports and using APA format.

Spring 2010   Teaching Assistant, Cognitive Processes, University of Louisville.
                     Conducted review sessions, guest lectures, and aided individual
                     students.

Spring 2009   Teaching Assistant, Experimental Psychology, University of
                     Louisville. Lab sections involved instructing students on aspects of
                     research design, writing research reports and using APA format.

Fall 2005   Teaching Assistant, C programming, Centre for Behavioural and
                     Cognitive Sciences, University of Allahabad
                     Taught computer programming (C Language).

Publications:

Chariker, J. H., Naaz, F., & Pani, J. (2011). Computer-based learning of neuroanatomy:
        A longitudinal study of learning, transfer, and retention. *Journal of Educational
        Psychology, 103(01)*, 19 – 31.

Chariker, J. H., Naaz, F., & Pani, J. (2012). Item difficulty in the evaluation of computer-
        based instruction: An example from neuroanatomy. *Anatomical Sciences
        Education, 5(2)*, 63-75.

Manuscripts:

Pani, J. R., Chariker, J. C., & Naaz, F. (2012). *Computer based instruction of neuroanatomy: Integrated learning of anatomical representations is more efficient than basic transfer of learning*. Manuscript submitted for publication.

Chariker, J. H., Naaz, F., & Pani, J. (2012). *Computer-based learning of neuroanatomy: Introduction to the neuroscience classroom*. Manuscript in preparation.

Presentations:

Naaz, F., & Srinivasan, N. (2005). *Effect of spatial cuing on change detection in natural scenes*. Paper presented at the 13th Annual Meeting of NAOP, Faizabad, India.

Naaz, F., & Srinivasan, N. (2006). Attention, awareness and knowledge: Implications of change blindness. In S. Menon (Ed.), *Consciousness, Experience and Ways of Knowing* (pp. 184 - 200). Bangalore, National Institute of Advanced Studies.

Chariker, J. H., Naaz, F., & Pani, J.R. (2009, November). *Computer-based learning of neuroanatomy: A longitudinal study of learning, transfer, and retention*. Paper presented at the 50th Annual Meeting of the Psychonomic Society, Boston, USA.

Chariker, J. H., Naaz, F., & Pani, J.R. (2010, July). *Computer-based learning of neuroanatomy*. Paper presented at 15th International Conference on Computer Games: AI, Animation, Mobile, Interactive Multimedia, Educational & Serious Games, Louisville, Kentucky, USA.

Chariker, J. H., Naaz, F., & Pani, J. R. (2010, August). *Computer-based learning of neuroanatomy: A longitudinal study of learning, transfer, and retention*. Paper presented at the Annual Meeting of the Cognitive Science Society, Portland, Oregon, USA.

Pani, J.R., Chariker, J. H., & Naaz, F. (2010, November). *Providing multiple representations for learning neuroanatomy produces transfer and benefits long-term retention*. Paper presented at the 51st Annual Meeting of the Psychonomic Society, St. Louis, Missouri, USA.

Chariker, J. H., Naaz, F., & Pani, J.R. (2011, November). *Large variations in item difficulty in computer based learning of neuroanatomy*. Paper presented at the 52nd Annual Meeting of the Psychonomic Society, Seattle, Washington, USA.