# Effect of Word Bank Provision for Lab Practicals on Student Performance in Human Anatomy and Physiology I and II Courses

**Carol A. Britson, PhD**
Department of Biology; University of Mississippi; University, MS 38677 cbritson@olemiss.edu

## Abstract
Providing word banks for human anatomy and physiology lab practicals is uncommon because instructors want to avoid cueing effects, but there is little published data on their effects on student performance. In the 2016-2017 academic year, word banks were not provided for students taking lab practicals in Anatomy and Physiology I and II while in 2017-2018 alphabetized word banks were provided. All other aspects of the courses remained the same. Student performance was significantly higher, though the effect size was small, on both lab practicals in Anatomy and Physiology I ($p<0.001$ for each) and the first lab practical in Anatomy and Physiology II ($p<0.001$). Scores on the 2nd lab practical in Anatomy and Physiology II and each lecture-lab course score were not affected by the provision of word banks. Results of this study will be useful when making decisions about providing word banks for lab practicals, weighting lab practicals for calculating grades, and responding to disability accommodation requests. https://doi.org/10.21692/haps.2020.004

**Key words:** lab practical, practical examination, word bank, anatomy, physiology, education

## Introduction
This study was developed in response to multiple student requests for a word bank to be provided during laboratory practical examinations in human anatomy and physiology courses. These requests were always answered with a simple 'no' and some verbiage about the importance of assessing factual recall of anatomic identification without any hints or clues that would artificially inflate a student's score and bias the effectiveness of the assessment. These answers were typically met with a bit of shock as students may have been expecting a word bank on a recall examination (Glass et al., 2007). Students were apprehensive due to the high level of test anxiety in individuals who are pursuing the health professions and must complete multiple high-stakes exams to complete their programs (Schwartz et al., 2015). Yet, these answers were provided based on personal experiences as a student, graduate teaching assistant, and as an instructor rather than on experimental data. In other words, 'we've always tested like this in the lab, so this way must be the right way.'

In a typical lab practical (or "spotter") exam, the focus is often solely on knowledge recall (Choudhury et al., 2016; Smith and McManus, 2015; Yaqinuddin et al., 2013). The words "recall" and "identify" are active learning verbs frequently used at the first level of learning (Knowledge) in the anatomy-specific modification to Bloom's Taxonomy, the Blooming Anatomy Tool (BAT) (Thompson and O'Loughlin 2015). While questions at this level are straight forward often with answers stated verbatim (Thompson and O'Loughlin 2015), correctly identifying anatomical structures enables the student to advance higher levels of learning. The importance of "identify" as a desired learning outcome can be highlighted by the number of mentions of the word "identify" in the HAPS Anatomy Learning Outcomes (108 mentions, HAPS 2019a) and HAPS Anatomy and Physiology Learning Outcomes (104 mentions, HAPS 2019b). Correct identification as a learning goal is appropriate in courses such as Human Anatomy and Physiology I and II which are often prerequisites (i.e., the foundation) to further study.

To assess the "identify" learning outcome in students of human anatomy and physiology, the traditional laboratory practical, often called "spotter" exam, typically utilizes a series of labeled specimens (i.e., stations), each with questions relating to the specimen which the students must answer in the given time before moving on to the next specimen (Choudhury et al., 2016). Though formats for lab practicals vary across instructors and institutions, published examples of formats include the following procedures: students bring only a writing instrument into the examination (Krippendorf et al., 2008), are not allowed to touch the specimens (Choudhury et al., 2016; Sagoo et al., 2016), write a free-text response on an answer sheet (Shaibah and van der Vleuten 2013), are given a set time (60 seconds, Krippendorf et al., 2008; 90 seconds, Sagoo et al., 2016) to answer the question(s), and move to the next question station (often laid out in a circular stream; Inuwa et al., 2012) at the end of the interval.

A variation of the spotter exam that moves the level of learning assessed to the Comprehension level (i.e., the question is straight forward, but the answer requires more than a simple definition) in Thompson and O'Loughlin's (2015) BAT is the steeplechase examination. Here, the student is asked to identify a structure and then name an associated

*continued on next page*

function or action (Smith and McManus, 2015). The objective structured practical examination (OSPE) pushes the level of learning assessed to the Application level of BAT (Thompson and O'Loughlin, 2015) by requiring that the student integrate anatomical knowledge with clinical skills (Choudhury et al., 2016; Yaqinuddin et al., 2013).

Schuwirth and van der Vleuten (2011) proposed the following four criteria to consider before deciding on an assessment format:

1. Is the assessment actually measuring what it is designed to measure (validity)?
2. Does the assessment produce consistent scores across different student cohorts (reliability)?
3. Does the assessment positively affect student learning and preparation (educational impact)?
4. Is the assessment a burden on instructional time (cost)?

The collective time required to develop questions for the practical, physically set up question stations, provide examination security by resetting the practical (e.g., switching questions between sections of students), marking answer sheets, and normalizing grading practices across instructors and teaching assistants is a cost to instructors. A cost to the institution is the functional limitation of the lab space during lab practical week as practical examinations are usually conducted on specimens in the anatomy laboratory (Inuwa et al., 2012).

Several investigators have developed protocols for lab practicals to minimize costs (typically of time and effort) while maintaining or enhancing validity, reliability, and educational impact of the lab practical. Manual grading of free text responses presents several costs and is prone to human error (e.g., correct answers may be marked wrong, incorrect answers may be given credit, total scores may be miscalculated; Krippendorf et al., 2008). Additionally, consistency in marking free text responses can be questioned (Choudhury et al., 2016), particularly across multiple instructors. Costs of time and inconsistency can be reduced by providing a choice of potential answers and using computer-graded assessments of lab practicals (Gentile et al., 2019), but they must be balanced by verifying that the assessment is valid. In other words, does providing a choice of potential answers for lab practical examinations accurately measure "correctly identify ___" learning outcomes, or are student scores distorted from the cueing effects (Damjanov et al., 1995; Fenderson et al., 1997) available from the list of choices?

Extended matching tests, where students select the best answer to a question from a list of 20 options, each of which may be used once, more than once, or not at all (Fenderson et al., 1997) are one possible alternative but there are far more than 20 options for a typical lab practical. Increasing the length of the options to several hundred or more alphabetized items creates an uncued exam (Fenderson et al., 1997). Krippendorf et al. (2008) created an uncued lab practical for first-year medical students by:

1. Providing students at the beginning of the course with lists of structures that would be on the lab practicals.
2. Providing students with a numbered, alphabetized list that they could refer to during the lab practicals.
3. Having students write the number of their selected item on their answer sheets.
4. Using an optical scanner to automatically grade the practical.

Krippendorf et al. (2008) found this method greatly reduced faculty grading time, reduced grading errors, and provided faster performance feedback for students without changing overall student performance. However, no student scores were reported nor were the results of any statistical tests performed (Krippendorf et al., 2008).

In a different study, Shaibah and van der Vleuten (2013) tested the performance of 100 gross anatomy students on a steeplechase lab practical in which answers could be provided via free text response or multiple-choice questions, each with five options. Average performance of students was significantly higher when multiple choice questions were used (91.17% $\pm$ 10.58 SD to 87.17% $\pm$10.84 SD, $p<0.001$). Shaibah and van der Vleuten (2013) raise the discrepancy between their results and those of Krippendorf et al. (2008) and suggest that the difference in results could be due to one of two reasons: 1. the populations used in the Krippendorf study were independent; or 2. cueing effects were removed by using a long list of about 300 options instead of using five options. Yet, the two studies cannot be compared because there are no data reported in Krippendorf et al. (2008).

Since there are no reported data supporting the presence or absence of the effect of word banks on student performance on lab practical examinations, there is no real mechanism to defend the position that providing a word bank fundamentally alters the nature of the lab practical examination by providing visual cues that provide prompts or hints that may artificially inflate a student's score. Thus, this study will test the null

continued on next page

hypothesis that providing word banks for lab practical examinations has no significant positive or negative effect on student performance on the following:

1. Lab practical examination #1.
2. Lab practical examination #2.
3. Four-credit lecture-lab courses in which lab practical examinations are part of the laboratory curriculum.
4. Hypothetical one-credit lab only courses.

The latter null hypothesis was tested because:

1. Course numbering systems for Anatomy and Physiology lecture and lab courses are unique to the academic institution.
2. The effect of providing word banks for lab practicals may be more impactful for instructors teaching one-credit laboratory courses in which 50% of a student's grade may be determined by their scores on lab practicals.

## Materials and Methods

*General Course Description and Student Profile*

At the University of Mississippi, Human Anatomy and Physiology I (hereafter referred to as A&P I) and Human Anatomy and Physiology II (hereafter referred to as A&P II) are general education, non-majors service courses offered through the Department of Biology. Other than admission to the university, there are no prerequisites to enroll in A&P I. Admission requirements did not change during this study. Students must earn a grade of C or better in A&P I to enroll in A&P II. Both courses are four-credit, lecture-lab courses in which the student receives a single letter grade for the course.

The content included in A&P I begins with an introduction to the human body and a review of chemistry and cells. The course continues with the study of tissues and the integumentary, skeletal, muscular, and nervous systems (excluding the special senses). In A&P II content begins with the special senses and continues with the endocrine, cardiovascular, immune, respiratory, digestive, urinary, and reproductive systems.

During a regular academic year, A&P I is only offered in the fall semesters, and A&P II is only offered in the spring semesters. Both A&P I and II are offered during summer sessions, but data from summer session performance is not included in the study.

The only difference in the administration of the A&P I and II courses between the 2016-2017 and 2017-2018 academic years was the provision of word banks during lab practicals in the 2017-2018 academic year.

For each course, there was a single lecture section meeting at 8am on Mondays, Wednesdays, and Fridays in the same auditorium (capacity of 394 students) at the University of Mississippi with the same instructor of record (Britson). The same editions of the lecture textbook (Amerman, 2016), lab manual (Whiting, 2016), and online resources (Pearson Education's Mastering™ A&P) were used throughout both academic years as was the university's course management system, BlackBoard™. Laboratory sections were capped at 30 students per section, staffed by two teaching assistants (TAs), held in the same laboratory room at the University of Mississippi, and met for two hours once per week beginning in the second week of the semester.

In A&P I there were 13 lab sections that met beginning at 11am on Tuesdays, Wednesdays, and Thursdays (four sections per day) with one section beginning at 11am on Fridays. In A&P II there were nine lab sections that met beginning at 11am on Tuesdays, and Wednesdays (three sections per day), 11am on Thursdays (two sections per day), and 1 section beginning at 11am on Fridays. There were nine TAs for A&P I in 2016 and 13 TAs in 2017 with four of the these being TAs in both years. There were nine TAs for A&P II in both 2017 and 2018 with four of the these being TAs in both years. All TAs were undergraduates that successfully passed both A&P I and A&P II with scores of 87% or higher and were selected through an identification and interview process (Hopp et al., 2019).

Additionally, there were two supplemental instruction (SI) leaders each semester that held a minimum of six (total), one-hour, peer-led study sessions each week of the course. Each SI leader served for both fall and spring semester within an academic year, and one of the SI leaders served in this role for both academic years. All SI leaders were undergraduates that successfully passed both A&P I and A&P II with scores of 80% or higher and were selected through an identification and interview process. Prior to beginning their time as an SI leader, they also underwent several training sessions administered through the university's Center for Excellence in Teaching and Learning.

For each course, grades were calculated and weighted from the following assessments: lecture exams (five total, 60% of grade), lecture quizzes (five total, 5% of grade), lab quizzes (ten total, 8% of grade), formative in-lab assessments (11 total, 7% of grade), lab practicals (two total, 10% of grade), and online homework (four to six per week, 10% of grade). Lecture exams were 50 multiple choice questions each; lecture quizzes were five multiple choice questions each; lab quizzes were ten questions based on the pre-lab readings and consisted of multiple choice, fill in the blank, or true-false questions; and in-lab assessments were short activities,

continued on next page

experiments, or round-robin quizzes. The round-robin quizzes and lab practicals will be described later in this section. Online homework consisted of 10-15 multiple choice questions per assessment (from both lecture and laboratory content) that were "open-book" assessments, but questions were randomly pulled from a pool of 40-50 questions and students were only allowed one attempt on the assessment. Except for the online homework, all assessments were completed in a face-to-face classroom setting. Each course was structured such that there was no more than one "high-stakes" assessment (e.g., lecture exam, lecture quiz, or lab practical) per week. Relative weights for each category of assessment were determined based on assessment difficulty and relative contribution of the lecture (75%) and laboratory (25%) contact hours.

The student group in 2016-2017 (word banks not provided) was comprised of 302 students enrolled in A&P I for the fall semester and 231 students enrolled in A&P II for the spring semester. The student group in 2017-2018 (word banks provided) was comprised of 318 students enrolled in A&P I for the fall semester and 200 students enrolled in A&P II for the spring semester. Declared majors of students enrolled in the courses were exercise science (30%), allied health including nursing and nutrition(28%), other science majors (e.g., biology and pharmacy, 28%), and other liberal arts or applied science majors (14%).

The demographic breakdown by major per year has been consistent since these data were collected beginning in 2011. Prior research on students enrolled in these two courses revealed that most students were interested in careers in nursing, physical therapy, occupational therapy, physician assistant, or dietetics (Hillhouse and Britson, 2018; O'Connor and Britson, 2017). Almost all students were in their 2nd or 3rd year of undergraduate education, were traditional college students between the ages of 18 and 23 years, and varied in race and gender. This study (Protocol #19x-003) was reviewed by The University of Mississippi's Institutional Review Board (IRB) and was approved as Exempt under 45 CFR 46.101(b) (#1 and 4).

*Laboratory Content and Lab Practical Format*

A weekly list of specific topics, the format of the in-class lab assessment, and the number of items on the weekly "Need to Know" list for A&P I and A&P II during both academic years is presented in Table 1. The Need to Know lists were posted on BlackBoard™ and were always available to the students. The in-class lab assessments were the only assessments in the course that were completed by a small group of three to four students (all other assessments were completed individually) and consisted of a variety of short activities (e.g., concept mapping, sensory receptor density, etc.), data acquisition system experiments (e.g., electromyography, blood pressure, etc.), and round robin quizzes. That latter of which were formative assessments designed to prepare students for the lab practicals and were very popular with the students.

To prepare for the round robin quizzes each group of students (a maximum of four per workstation, eight total workstations) were given a short selection of six to eight items from the weekly Need to Know list. They were tasked with using the models and specimens available at their workstation to create two identification questions (e.g., identify this tissue, identify this bone, etc.) for the round robin quiz. With all groups contributing, a short quiz of 16 identification questions was produced. The round robin quiz began when there was 15 minutes remaining in the lab session. Each group of students could bring the answer sheet and one copy of the Need to Know list with them as they walked around to each workstation, allowed two minutes at each workstation, conferred within their group, and answered the available questions. At the end of the quiz, each group submitted their answers and copy of the Need to Know list.

Lab practicals were administered twice each semester during weeks seven and 14 of a 15-week term. A word bank was prepared for each lab practical by combining and alphabetizing all the items on the Need to Know lists preceding each lab practical. Items were alphabetized to prevent cueing effects (Damjanov et al.,1995; Fenderson et al., 1997) that may have been present if items were grouped by week or body system. A small subsample of ten items from each word bank are presented in Table 2.

Each lab practical consisted of 50 questions that were presented at 16 "stations" throughout the laboratory classroom, with three to four questions per station. The number of questions per topic was determined by identifying the total number of items from the combined Need to Know list, dividing the number of each week's Need to Know items by the total, and then multiplying the result by 50. For example, there were 341 total items for the second lab practical of A&P I. There were 32 items for week nin e content which represented 9.4% of the total number of items. After multiplication and rounding, five of the 50 questions on the second lab practical were devoted to muscle tissue and an introduction to the muscular system.

Week eight had a much larger percentage of the total content (109 items representing 31.9%) and 16 questions on the lab practical were devoted to the appendicular skeleton and joints. This proportional division of the 50 questions was performed to reflect the relative difficulty and amount of time students would spend studying each area of content in preparation for the lab practicals. This proportional division of questions is similar to the matrix method used by Smith and McManus (2015) to determine the number of questions per content as based on the number of course hours per topic.

continued on next page

| | A&P I | | | A&P II | | |
|---|---|---|---|---|---|---|
| **Week** | **Topic** | **In-class Assessment** | **NTK list items** | **Topic** | **In-class Assessment** | **NTK list items** |
| 1 | no labs | | | no labs | | |
| 2 | Intro. to A&P; Intro to Organ Systems | Study strategies worksheet | 112 | General Senses | Sensory receptor density | 17 |
| 3 | The Microscope; The Cell | Memory matrix | 33 | Special Senses #1 | Round robin quiz | 61 |
| 4 | Histology | Concept mapping | 39 | Special Senses #2; Endocrine System | Diabetes experiment | 34 |
| 5 | Integumentary System; Intro to Skeletal System | Round robin quiz | 70 | Blood | Blood assays | 15 |
| 6 | Axial Skeleton | Round robin quiz | 90 | Anatomy and Physiology of the Heart | ECG experiment | 42 |
| 7 | Lab Practical #1 | | | Lab Practical #1 | | |
| 8 | Appendicular Skeleton; Joints | Round robin quiz | 109 | Blood Pressure; Blood Vessels #1 | BP experiment | 21 |
| 9 | Intro to Muscular System: Muscle Tissue | Intro to data acquisition equipment | 32 | Blood Vessels #2; Lymphatic System | Round robin quiz | 68 |
| 10 | Muscular System | EMG experiment | 71 | Anatomy of Respiratory System | Round robin quiz | 47 |
| 11 | Intro to Nervous System | Round robin quiz | 35 | Physiology of Resp. System; Anatomy of Digestive System | Lung volumes experiment | 69 |
| 12 | Central Nervous System: Brain and Spinal Cord | Round robin quiz | 53 | Anatomy and Physiology of the Urinary System | Round robin quiz | 26 |
| 13 | Peripheral Nervous System: Nerves | Reflex experiment | 41 | Reproductive Systems; Development | Round robin quiz | 68 |
| 14 | Lab Practical #2 | | | Lab Practical #2 | | |

***Table 1.*** *Weekly list of specific topics, the format of the in-class lab assessment, and the number of items on the weekly "Need to Know" list for A&P I and A&P II during the 2016-2017 and 2017-2018 academic years.*

continued on next page

All lab practical questions were developed by the instructor of record (Britson), and the physical question stations were prepared by the laboratory TAs. The development of questions included preparing for "resets" of answers each day as a different answer key was used for each day's lab section. Resetting a lab practical consisted of moving a label, switching a microscope slide, etc. For example, on the first day of the lab practical, the question may be 'Identify this sublayer' with the answer of 'stratum lucidum'. On the second day of the lab practical, the answer may be 'stratum basale', on the third day, the answer may be 'stratum spinosum', and so on. Over the course of four days of testing, the majority of the items were used as an answer item.

The 30 students enrolled in each lab section were divided into two groups to take the lab practical during the first or second hour of the lab session. Upon entering the lab, they were given a numbered answer sheet with the word bank attached. Students were given approximately three minutes per station and only one student was allowed per station. At the end of a full rotation, students were given five minutes to revisit any station they wished before submitting their answer sheet with the word bank still attached. Two TAs supervised testing at all times. A separate extended testing session was provided for students with approved testing accommodations.

Before entering scores in the Learning Management System, all TAs met with the instructor or record (Britson) for grade normalization, a procedure which has been used since 2014 in maintain reliability and validity of grading procedures and scores across sections and years. During this meeting, the instructor of record made all final decisions about answers that would receive full credit (two points), partial credit (one point), or no credit (zero points). As an example, if the correct answer was "descending colon", an answer of ascending colon would receive partial credit. All students were held to the same policy on spelling (i.e., two incorrect letters allowed as long as the meaning of the term is not affected and 0.5 points off for each additional incorrect letter). Students with approved spelling accommodations are given an additional incorrect letter (i.e., three letters) before the deductions begin.

| **A&P I (fall semester)** | | **A&P II (spring semester)** | |
|---|---|---|---|
| **LP1** | **LP2** | **LP1** | **LP2** |
| Cardiac muscle | Basal nuclei | Chordae tendineae | Capillary |
| Cardiovascular | Biceps brachii | Choroid | Capsular space |
| Carotid canal | Biceps femoris | Ciliary body | Cardia of stomach |
| Carpal | Bipolar neuron | Circumflex artery | Cardiac notch |
| Central canal | Brachial plexus | Cochlea | Carina |
| Centriole | Brachialis | Cochlear branch | Cartilaginous rings |
| Centrosome | Brachioradialis | Cone | Cecum |
| Cephalic | Brain | Cornea | Central vein of liver |
| Cervical | Brainstem | Coronary sinus | Cephalic vein |
| Cervical curvature | Buccinator | Delta cell | Cervix of uterus |

***Table 2.*** *Subsample of 10 items from the combined Need to Know lists (i.e., word banks) provided for each lab practical (LP1 and LP2) in the 2017-2018 academic year in which word banks were provided. All items in each work bank were alphabetized, printed, and stapled to the answer sheet given to each student. The subsample was selected from approximately the same location (i.e., starting with the 30$^{th}$ item) in each list.*

continued on next page

*Analytical Methods*

For A&P I and II a one-way analysis of variance (ANOVA) with the level of significance set at alpha=0.05 was used to test the effect of providing word banks on student performance on the first and second lab practicals, the overall (4-credit, lecture and lab) course score, and a hypothetical lab-only course score. Relative contributions of online homework, lab quizzes, and formative in-lab assessments were maintained. For example, in a four-credit lecture-lab course lab practicals (two total) were weighted as 10% of the entire course grade but were 28.5% of the hypothetical lab-only course score. Lab quizzes (ten total), formative in-lab assessments (11 total), and lab-related only online homework assessments (22 total) were weighted as 22.9%, 20%, and 28.5% of the hypothetical lab-only course score, respectively. Effect sizes were calculated using Cohen's d statistic (McLeod 2019) for each one-way analysis. All analyses were performed using SPSS statistical package, Version 22 (SPSS, Chicago, IL) licensed to the University of Mississippi.

## Results

Student performance on both lab practicals was significantly higher when word banks were provided in A&P I [(F = 11.313; df = 1,619; p = 0.001 for Lab Practical 1), (F = 16.013; df =1,619; p < 0.001 for Lab Practical 2) (Figure 1)]. Overall numerical performance of students in a four-credit, lecture-lab A&P I course was not significantly different when word banks were provided for lab practicals (F = 3.601; df =1,619; p = 0.058; Fig. 1). For a hypothetical, one-credit lab-only A&P I course, student performance was significantly higher when word banks were provided (F = 8.214; df =1,619; p = 0.004; Figure 1). Effect sizes (d) for A&P I data were 0.268 for the first lab practical, 0.317 for the second lab practical, -0.152 for the four-credit, lecture-lab course score, and 0.228 for the hypothetical, one-credit lab-only course score. A Cohen's d value of 1 would indicate that means differ by 1 standard devation, a d of 2 indicates that the mean differ by 2 standard deviations, etc. (McLeod 2019). A d of 0.1 to 0.3 is considered a small effect size, a value of 0.3 to 0.5 a medium effect, and value above 0.5 a large effect (Cohen 1988).

In A&P II, student performance on the first lab practical was significantly higher (F = 35.194; df = 1,430; p < 0.001; Figure 2) when word banks were provided but was not significantly different for the second lab practical (F = 2.344; df =1,430; p = 0.126; Figure 2). Overall numerical performance of students in a four-credit, lecture-lab A&P II course was not significantly different when word banks were provided for lab practicals (F = 0.280; df =1,430; p = 0.597; Figure 2). For a hypothetical, one-credit lab-only A&P II course, student performance was significantly

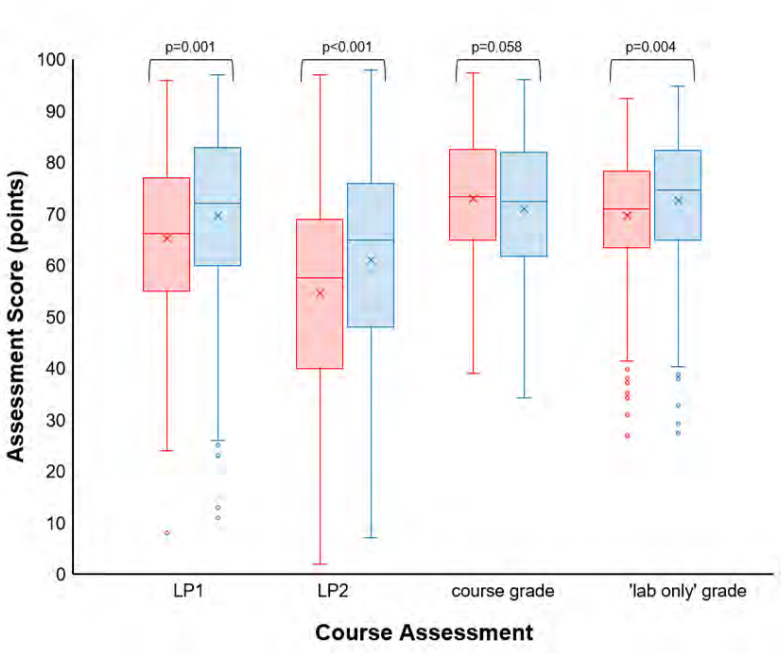


***Figure 1.*** *Human Anatomy & Physiology I scores at the University of Mississippi for Fall 2016 (red boxes, no word banks provided with lab practicals; n=302) and Fall 2017 (blue boxes, word banks provided; n=318). Scores for two lab practicals (LP1, LP2), numerical scores for a 4-credit lecture/lab course, and a hypothetical 1-credit lab only course are depicted. Within each box the assessment mean is represented by "x", the middle line is the assessment median, the bottom and top lines of the box represent the 1st and 3rd quartiles respectively, and the top and bottom bars represent the maximum and minimum values respectively. Individual data points (open circles) greater than 1.5 times the interquartile range represent outliers beyond the maximum or minimum values. ANOVA test results for between-subject effects are shown with exact p-values.*

continued on next page

higher when word banks were provided (F = 17.771; df =1,430; p < 0.001; Figure 2). Effect sizes (d) for A&P II data were 0.551 for the first lab practical, 0.147 for the second lab practical, -0.051 for the four-credit, lecture-lab course score, and 0.399 for the hypothetical, one-credit lab-only course score.

## Discussion

In this paper, the impact of providing word banks with lab practicals on student performance and course outcomes with a study that provides multiple, useful outcomes was examined. In A&P I, student scores were significantly higher on both lab practicals in the semester where word banks were provided, but the impact of large standard deviations relative to actual differences between the means (i.e., a small effect size) indicates that statistical significance was due to a large sample size rather than true, artificial inflation of scores.

In terms of classroom significance, as compared to statistical significance, the difference in means between the semester where word banks were provided and the semester where word banks were not provided is no more than two questions (out of 50) for the first lab practical and 3.5 questions for the second lab practical. For the first lab practical in A&P II, student scores were significantly higher in the semester where word banks were provided, but the effect size was stronger. The difference in mean scores was the point value of 3.5 questions and similar to student performance for the second lab practical in A&P I, but smaller standard deviations within the sample led to the increased effect size.

The decrease in magnitude of the standard deviations is likely a result of A&P II students having met the prerequisite of successfully passing A&P I with a grade of "C" or better. This prerequisite aligns with the HAPS (2019c) suggested, required prerequisite for A&P II courses. Student scores for the second lab practical in A&P II were not statistically different and represented a classroom difference of one more question correctly answered when word banks were provided. Throughout the 2017-2018 academic year teaching assistants communicated personal observations that students were making fewer spelling errors as compared to the 2016-2017 academic year, but students continued to leave some questions unanswered.

Shaibah and van der Vleuten (2013) observed significantly higher scores for students taking a lab practical with multiple-choice questions but questioned the impact of their results when compared to those of Krippendorf et al. (2008). Shaibah and van der Vleuten (2013) incorrectly inferred that there were no statistically significant

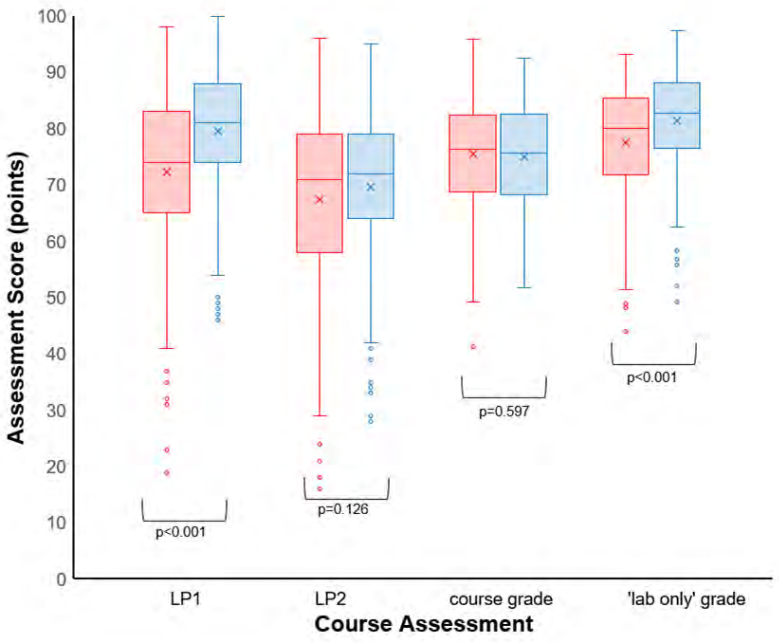


***Figure 2.*** *Human Anatomy & Physiology II scores at the University of Mississippi for Spring 2017 (red boxes, no word banks provided with lab practicals; n=231) and Spring 2018 (blue boxes, word banks provided; n=200). Scores for two lab practicals (LP1, LP2), numerical scores for a 4-credit lecture/lab course, and a hypothetical 1-credit lab only course are depicted. Within each box the assessment mean is represented by "x", the middle line is the assessment median, the bottom and top lines of the box represent the 1st and 3rd quartiles respectively, and the top and bottom bars represent the maximum and minimum values respectively. Individual data points (open circles) greater than 1.5 times the interquartile range represent outliers beyond the maximum or minimum values. ANOVA test results for between-subject effects are shown with exact p-values.*

continued on next page

differences in Krippendorf et al.'s (2008) paper. However, only observations that student scores were not "noticeably different" when students were given a list of terms during a lab practical rather than student performance data were reported in Krippendorf et al. (2008). With a sample size of 100 students taking a single lab practical successively in two different formats, Shaibah and van der Vleuten (2013) document a four-point increase in mean performance. While they did not report effect size, means and standard deviations for recall questions (24 total) reported in Table 2 of Shaibah and van der Vleuten (2013) can be used to estimate a small-to-moderate (McLeod 2019) effect size of 0.38. In their study, the difference in student performance is three more correctly answered questions in the multiple-choice lab practical.

Across the multiple assessments of an entire course, the impact of any one or two assessments was low. A philosophy central to the relative weightings of assessment categories for the courses in this study is 'no one assessment can make or break' a student's score for the course. The lack of significant differences in student performance in A&P I and II courses where words banks were not provided for lab practicals (2016-2017 academic year) and where word banks were provided (2017-2018 academic year) is reflective of the number and relative weights of each type of assessment for calculation of a student's course performance.

This study is not unique in analyzing student performance across multiple cohorts of students and different formats of lab practicals over entire courses. Choudhury et al. (2016) and Smith and McManus (2014) studied lab practical performance from six, year-long cohorts of optometry students and five, year-long cohorts of Bachelor of Medicine students, respectively in the United Kingdom. The present study, however, focuses on introductory, human anatomy and physiology students often in their first or second year of post-secondary education rather than upper division or professional school students.

To compare student performance in hypothetical, one-credit laboratory courses, the relative weightings of online homework supporting laboratory learning outcomes, lab quizzes, formative in-lab activities, and lab practicals were maintained for calculation of a lab course score. These hypothetical, one-credit laboratory course scores were significantly higher for A&P I and II when word banks were provided for students taking lab practicals. Though the effect size is small to medium in each comparison, a small numerical difference can have a larger, relative impact on a student's letter grade. In the calculation of a student's (or applicant's) grade point average, however, the "weighting" of a course grade is reflected in the credit hours earned. Thus, the effect of providing word banks for lab practicals in a one-credit course, though positively significant, would be unlikely to carry over to statistically significant effects, and large effect sizes, on the student's semester or overall GPA. A thorough study of the effects of relative weighting of assessments on student performance in human A&P courses is warranted, however, as McDonald et al. (2016) found that increasing the grade weighting of lab practicals, while maintaining all other aspects of curriculum design, in a human anatomy course led to improved scores on the lab practical assessments, the number of students passing the practical assessments, and performance on subsequent assessments in a four year study.

## Conclusions

Laboratory practical examinations provide instructors with the ability to assess students on their understanding of three-dimensional spatial relations of anatomical structures to others and their ability to differentiate between similar structures (Smith and McManus 2015). Despite the costs of time and resources to set up, deliver, and grade laboratory practicals (Schuwirth and van der Vleuten 2011), the educational value they provide motivates instructors to develop innovations that minimize costs.

Alternatives to the spotter format such as the steeplechase (Smith and McManus 2015) and objective-structured practical examination (Choudhury et al., 2016; Yaqinuddin et al., 2013) increase the educational impact of the lab practical assessment. Modifications to the set-up, delivery (such as the provision of word banks), and grading minimize costs but these reductions must not be at the expense of the validity and reliability of the assessment. This study has shown providing word banks for students taking lab practicals in human A&P I and II courses (1) increases student performance on the practicals though the effect size is small and (2) the overall score in a four-credit, lecture-lab course is unaffected. For the latter, careful consideration of how assessments are weighted for the calculation of a student's overall score is critical in judging the effect of providing word banks.

## Acknowledgements

I thank my current and former department chairs, G Roman and P Lago, as well as the Department of Biology at the University of Mississippi for supporting the development of the human anatomy and physiology laboratory and instructional program. G Roman's comments on the manuscript are also appreciated. M Bland is a phenomenal publisher representative, and her support and assistance have enabled me to build the instructional program to what it is today. Lastly, I thank the awesome teaching assistants (A Lawson, A Ramsey, BG Elkin, C Hancock, D Raines, E Landers, M Eubanks, M Whitehead, P Dunn, D Shands, N Francis, L Kraft, K Shetley, ML Hodge, M Marquez, A Eftink, MP Davidson) and supplemental instruction leaders (K Downie, K Brown, C Hennig) that I have worked with and relied on during the years described in this study.

continued on next page

## About the Author

Carol A. Britson, Ph.D., is an Instructional Associate Professor and Associate Chair for Undergraduate Studies in the Department of Biology at the University of Mississippi. She teaches Human Anatomy; Histology; Physiology; and Human Anatomy and Physiology I and II.

## Literature Cited

Amerman EC. 2016. Human Anatomy & Physiology, 1st Ed. Pearson Education, Inc.

Choudhury B, Gouldsborough I, and Shaw FL. 2016. The intelligent anatomy spotter: A new approach to incorporate higher levels of Bloom's taxonomy. *Anat Sci Ed* 9:440-445.

Cohen J. 1988. Statistical Power Analysis for the Behavioral Sciences, 2nd Ed. Routledge.

Damjanov I, Fenderson BA, Veloski JJ, and Rubin E. 1995. Testing of medical students with open-ended, uncued questions. *Hum Pathol* 26:362-365.

Fenderson BA, Damjanov I, Robeson MR, Veloski JJ, and Rubin E. 1997. The virtues of extended matching and uncued tests as alternatives to multiple choice questions. *Hum Pathol* 28:526-532.

Gentile M, Melovitz-Vasan C, Huff S and Vasan N. 2019. The utilization of ExamSoft-iPad technology in administering and grading anatomy practical examinations. *Med Sci. Educ* 29:831-840. https://doi.org/10.1007/s40670-019-00750-0

Glass LA, Clause CB, and Kreiner DS. 2007. Effect of test-expectancy and word bank availability on test performance. *College Student Journal* 41(2).

HAPS. 2019a. Human Anatomy and Physiology Society, Anatomy Learning Outcomes, Revised Fall 2019.

HAPS 2019b. Human Anatomy and Physiology Society, Anatomy and Physiology Learning Outcomes, Revised Fall 2019.

HAPS 2019c. Human Anatomy and Physiology Society, Course Guidelines for Undergraduate Instruction in Human Anatomy and Physiology. https://www.hapsweb.org/page/course_guidelines (accessed 8 February 2020).

Hillhouse KC and Britson CA. 2018. Bring your own device initiative to improve engagement and performance in Human Anatomy and Physiology I and II laboratories. *HAPS Educator* 22(1):40-49. https://doi:10.21692/haps.2018.004

Hopp R, Britson C, Mukhopadhyay S, Goldina A, Chapman M, and Nielsen M. 2019. Teaching assistants in human anatomy and physiology: Their prevalence, recruitment strategies, funding, retention, and training. *HAPS Educator* 23 (2): 367-84. https://doi.org/10.21692/haps.2019.015

Inuwa IM, Taranikanti V, Al-Rawahy M, and Habbal O. 2012. Anatomy practical examinations: How does student performance on computerized evaluation compare with the traditional format? *Anat Sci Ed* 5:27-32.

Krippendorf BB, Bolender DL, and Kolesari GL. 2008. Computerized grading of anatomy laboratory practical examinations. *Anat Sci Ed* 1:220-223.

McDonald AC, Chan SP and Schuijers JA. 2016. Practical session assessments in human anatomy: Weightings and performance. *Anat Sci Ed* 9:330–336.

McLeod SA. 2019 (July 10). *What does effect size tell you? Simply Psychology*. https://www.simplypsychology.org/effect-size.html

O'Connor AE and Britson CA. 2017. Analysis of an arthritis simulation activity developed as a laboratory exercise for allied health students. *HAPS Educator* 21: 30-38. https://doi:10.21692/haps.2017.005

Sagoo MG, Smith CF, and Gosden E. 2016. Assessment of anatomical knowledge by practical examinations: The effect of question design on student performance. *Anat Sci Educ* 9:446–452.

Schuwirth LW and van der Vleuten CP. 2011. General overview of the theories used in assessment: AMEE Guide No. 57. *Med Teach* 33:783–797.

Schwartz SM, Evans C, and Agur, AMR. 2015. Comparison of physical therapy anatomy performance and anxiety scores in timed and untimed practical tests. *Anat Sci Educ* 8:518-524.

Shaibah HS and van der Vleuten CPM. 2013. The validity of multiple choice practical examinations as an alternative to traditional free response examination formats in gross anatomy. *Anat Sci Educ* 6:149–156.

Smith CF and McManus B. 2015. The integrated anatomy practical paper: A robust assessment method for anatomy education today. *Anat Sci Educ* 8:63–73.

Thompson AR and O'Loughlin VD. 2015. The Blooming Anatomy Tool (BAT): A discipline-specific rubric for utilizing Bloom's Taxonomy in the design and evaluation of assessments in the anatomical sciences. *Anat Sci Educ* 8:493–501.

Whiting CC. 2016. Human Anatomy & Physiology Laboratory Manual: Making Connections, Main Version, 1st Ed. Pearson Education, Inc.

Yaqinuddin A, Zafar M, Ikram MF, Ganguly P. 2013. What is an objective structured practical examination in anatomy? *Anat Sci Educ* 6:125–133.

Back to TOC